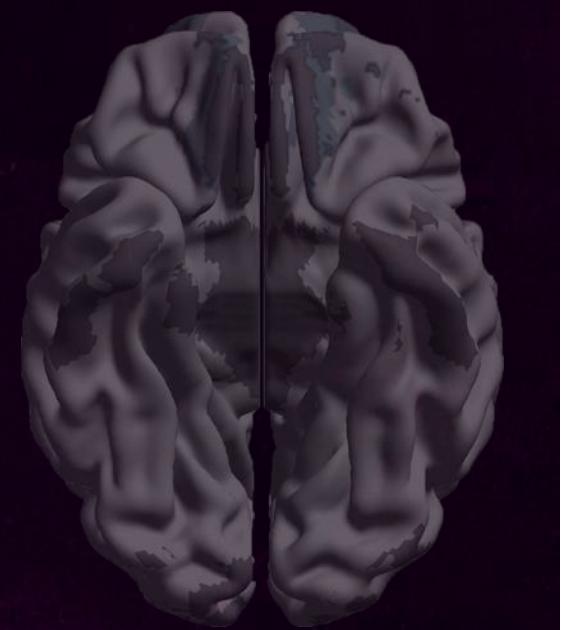


Estimación de la edad cerebral mediante registros de electroencefalografía (EEG) utilizando técnicas de aprendizaje de máquinas



Felipe Carriel Rubilar

Profesora Guía: María José Escobar

Profesor Correferente: Matías Zañartu

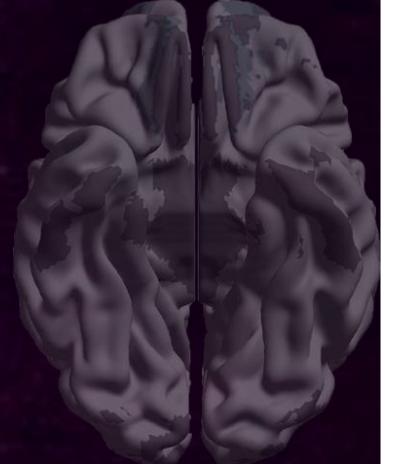
Profesora Coguía: Mónica Otero

Profesor Coguía: Pavel Prado



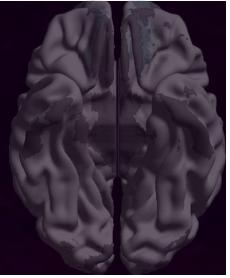
DEPARTAMENTO DE
ELECTRONICA

Objetivos

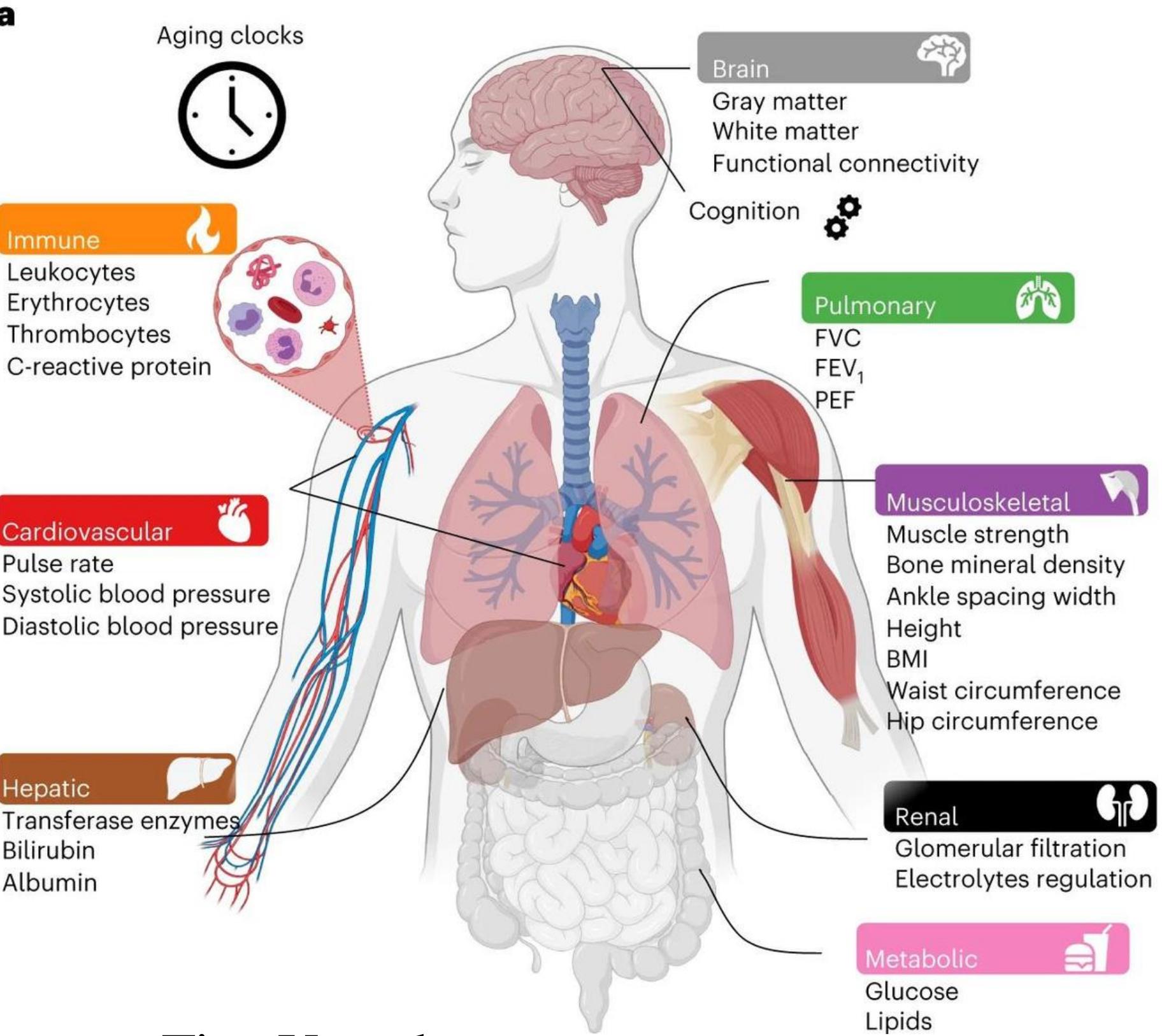
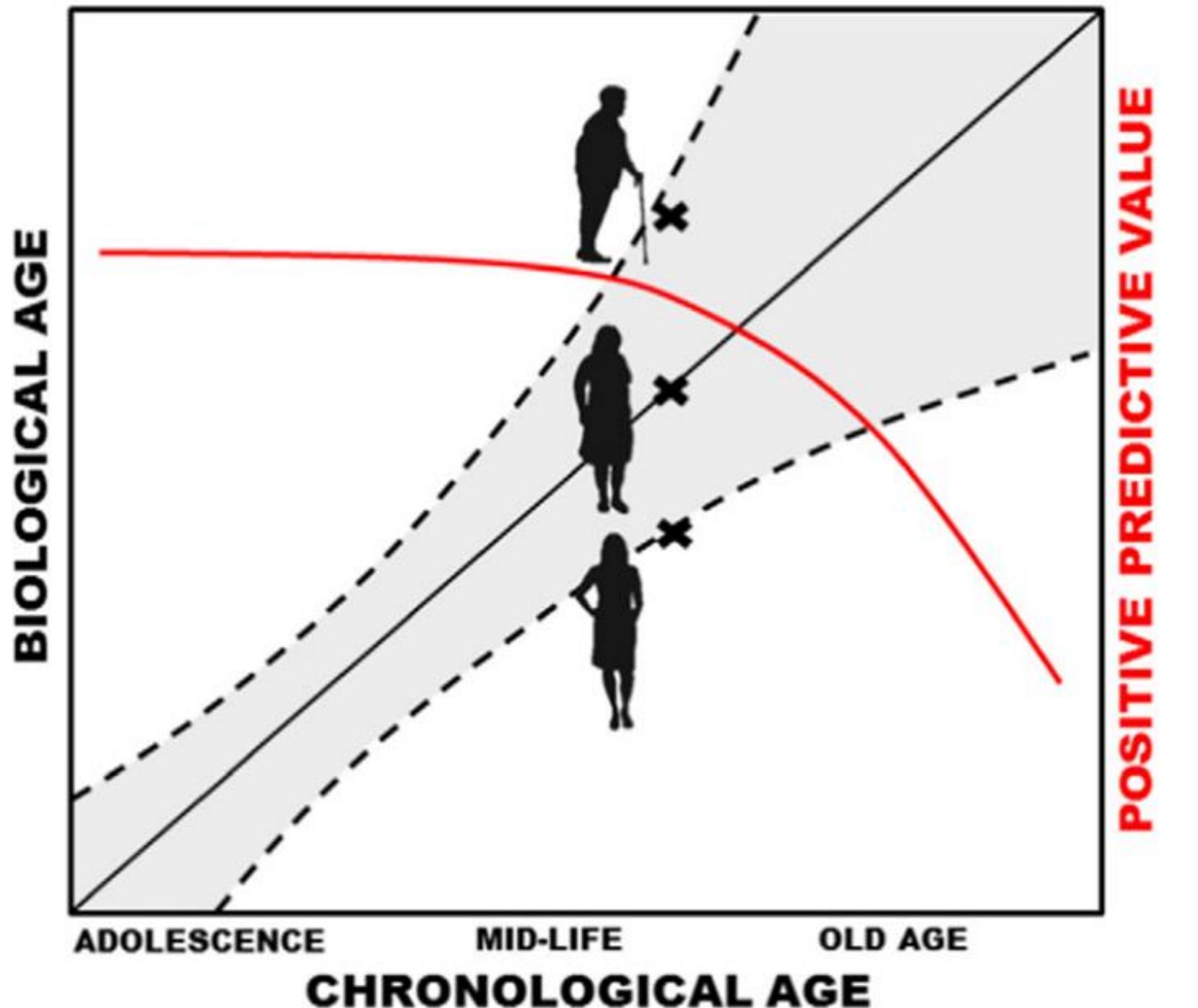


- Estimación de la edad biológica cerebral vs la edad cronológica.
- Evaluación de los distintos algoritmos de aprendizaje de máquina para estimación de la edad biológica cerebral.
- Identificar las características más relevantes en la estimación de edad cerebral.

Introducción

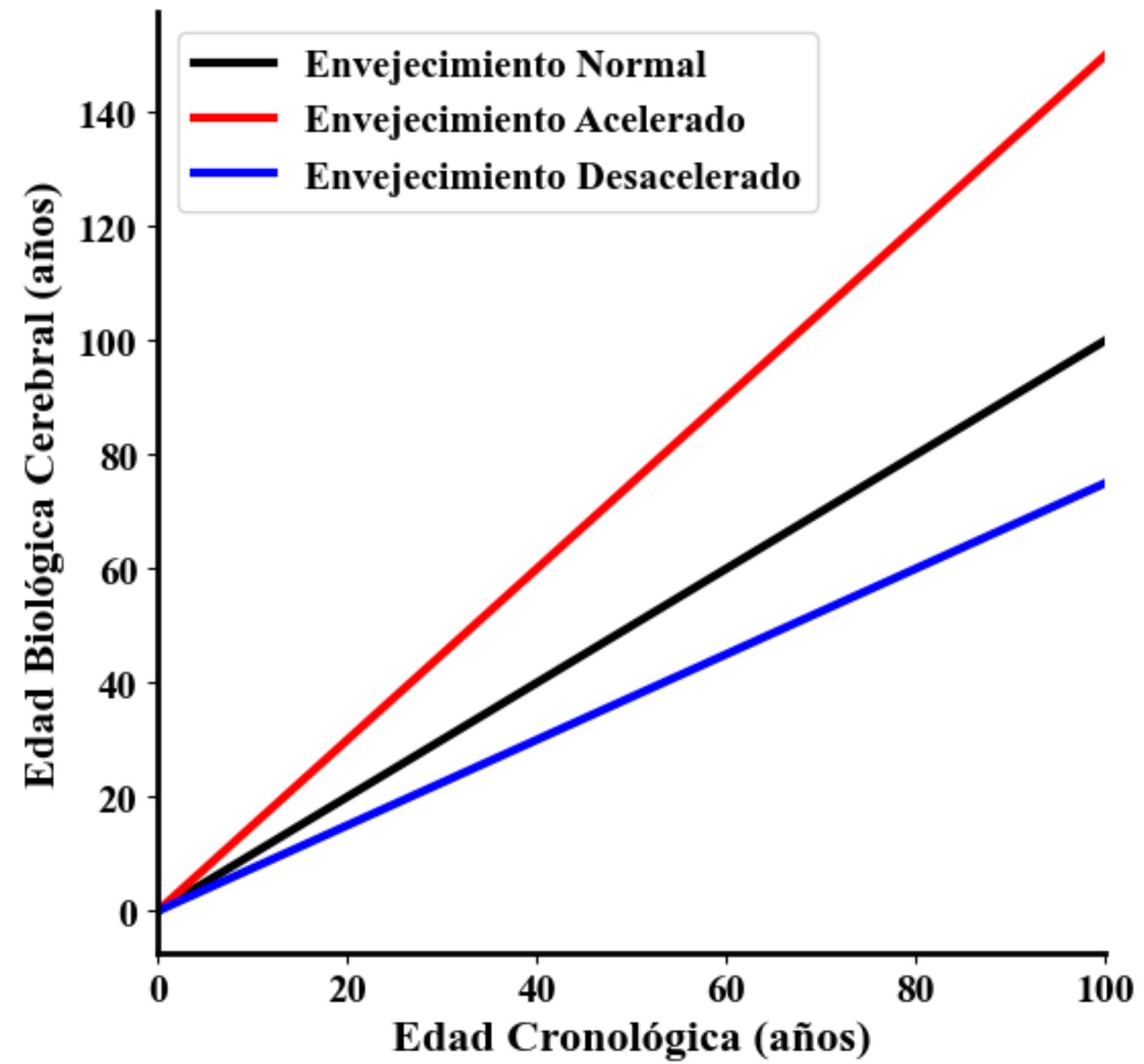
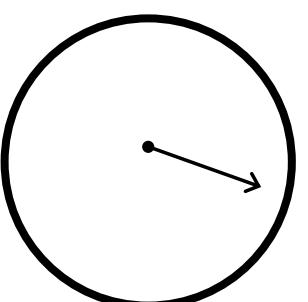
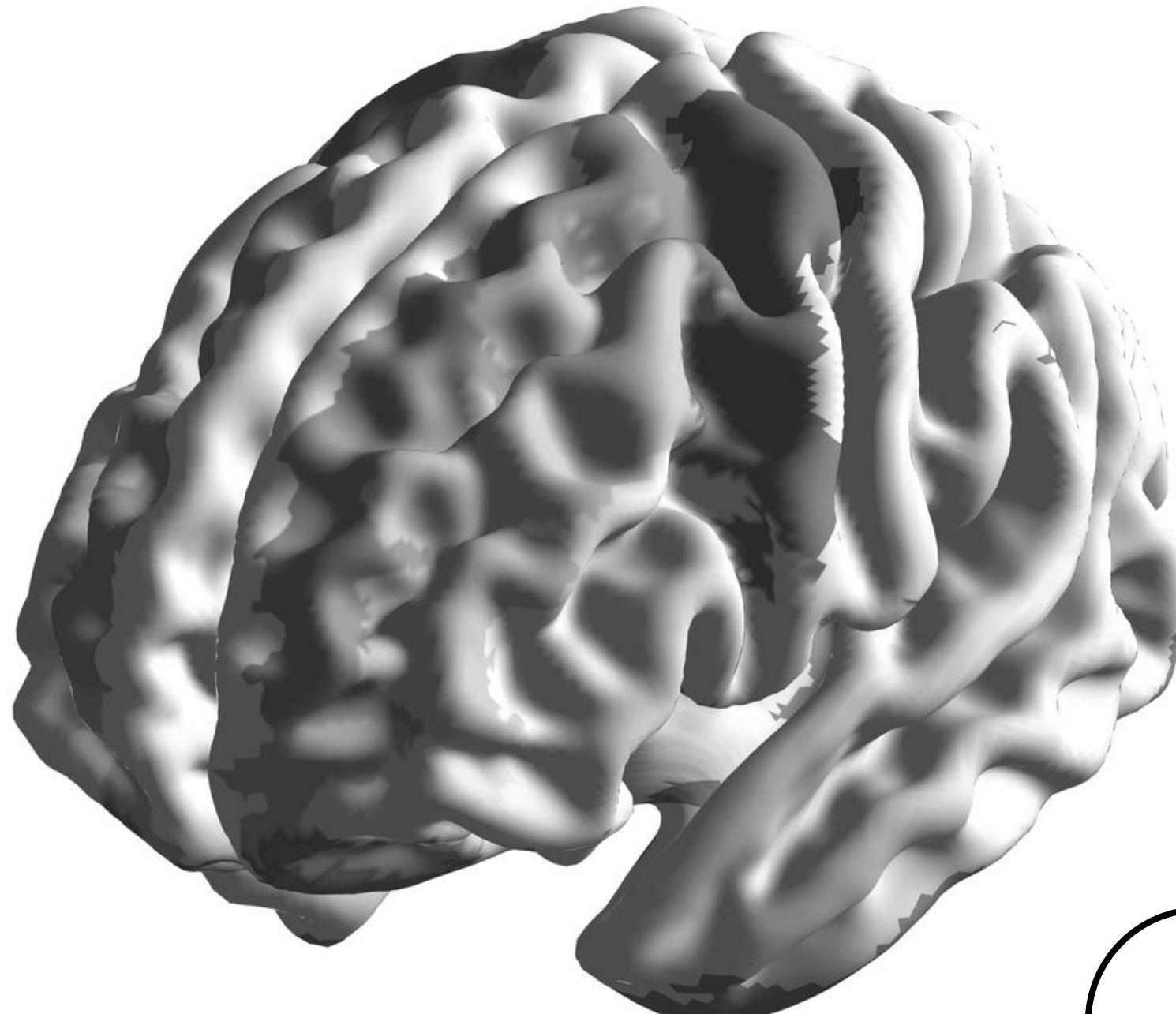
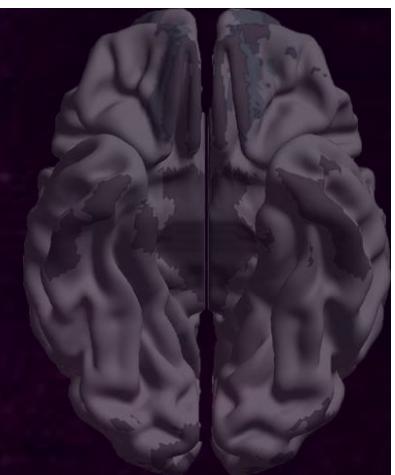


¿Qué es la edad biológica?

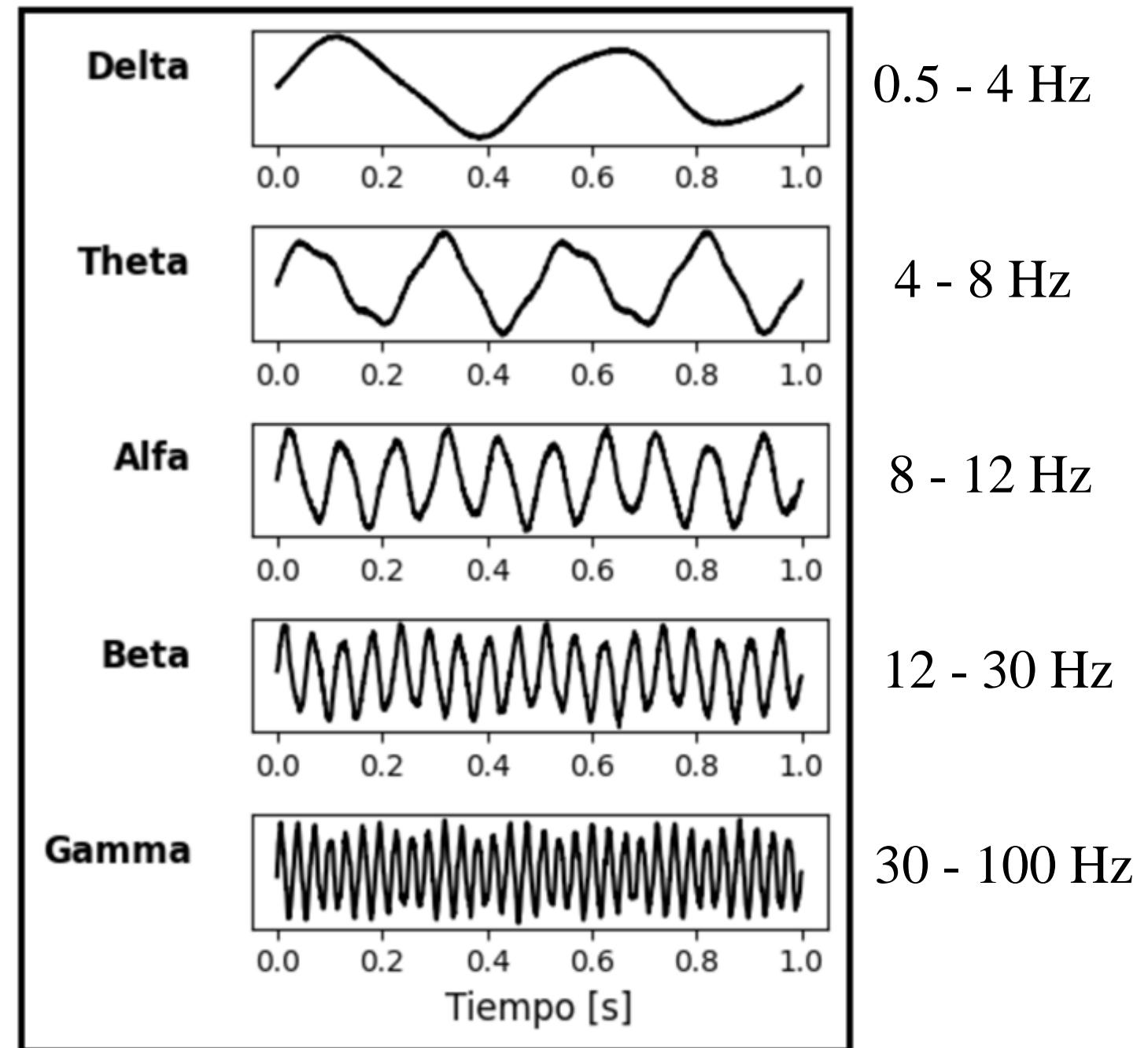
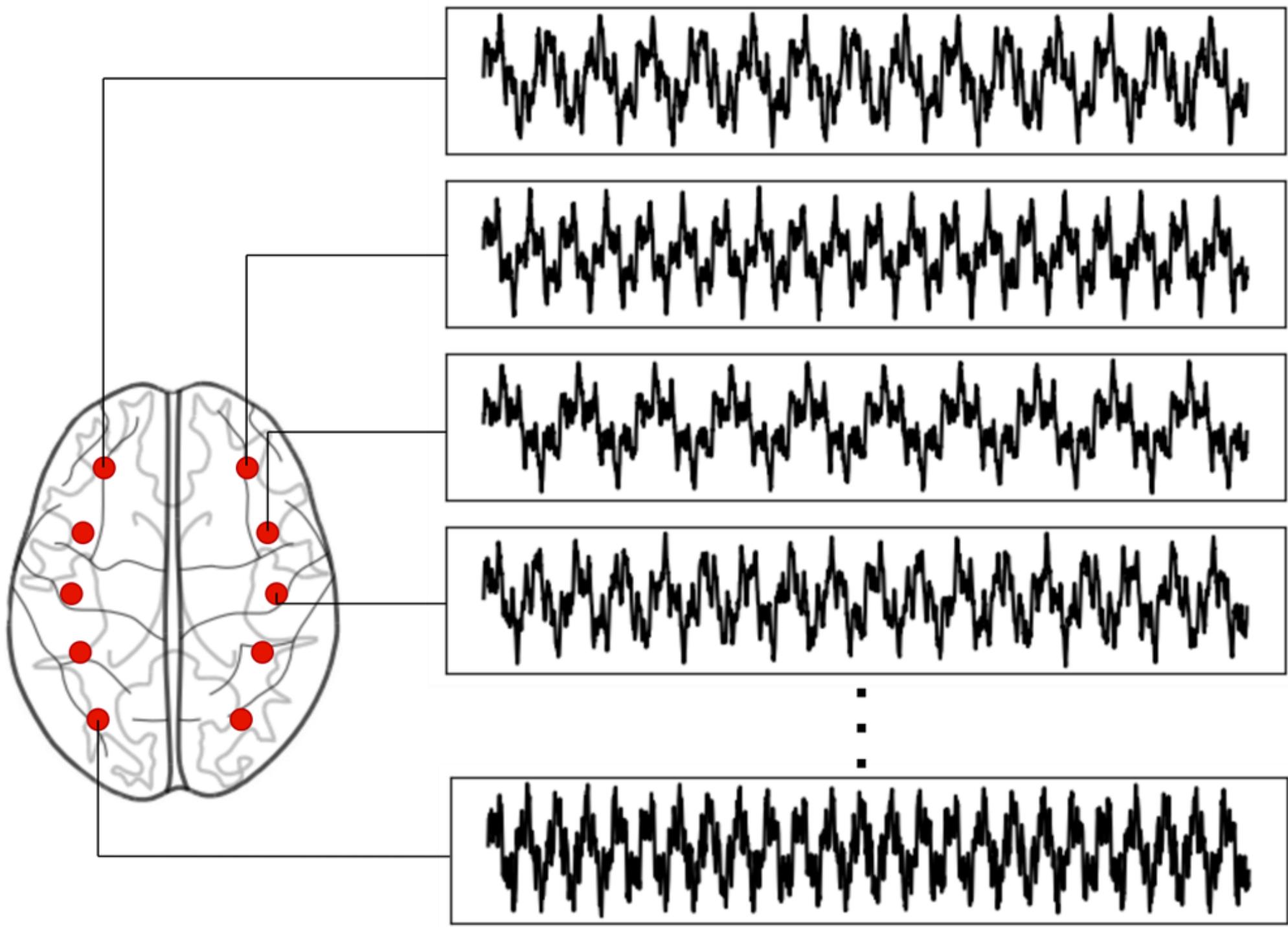
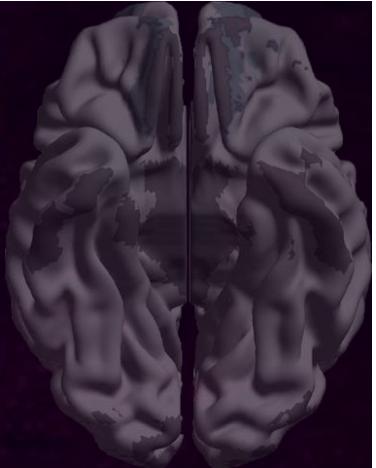


Tian Y et al.

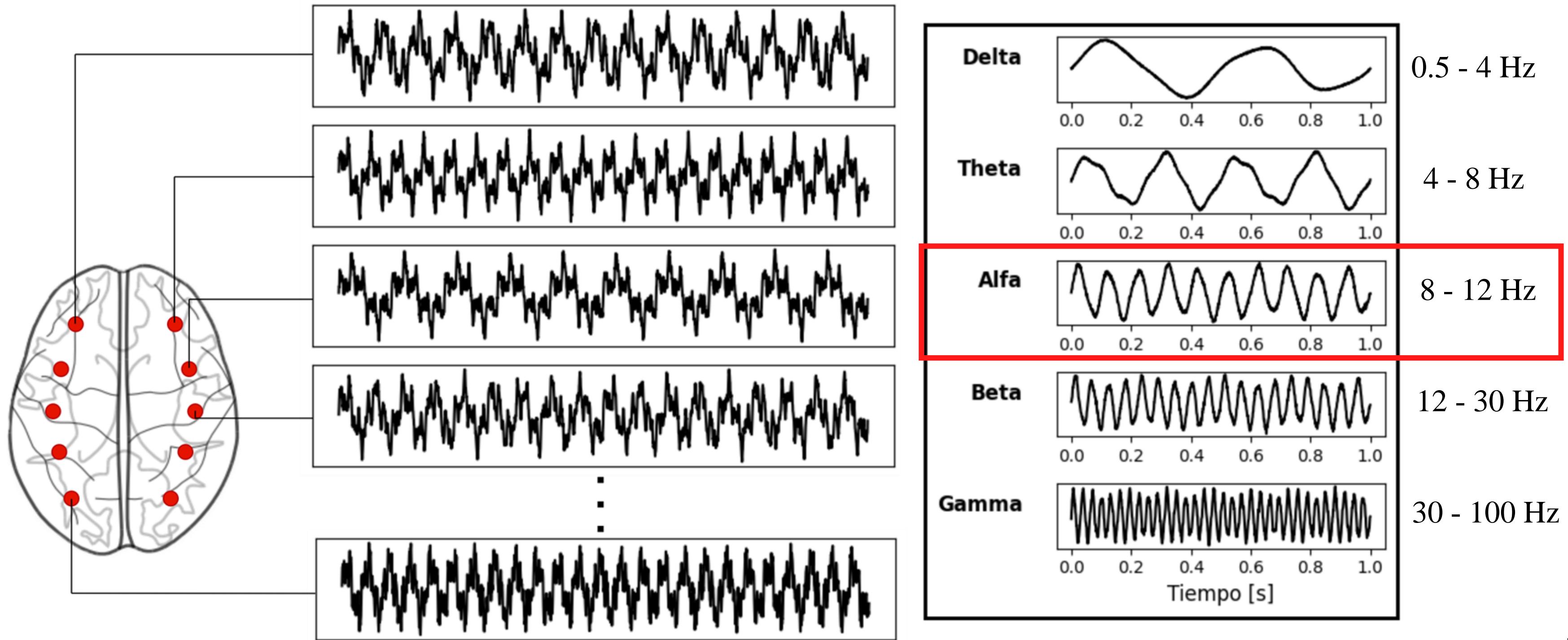
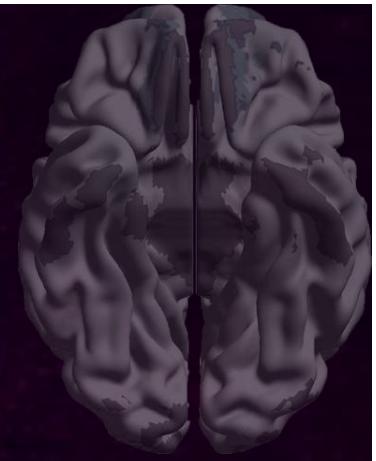
Edad Biológica Cerebral



Electroencefalograma

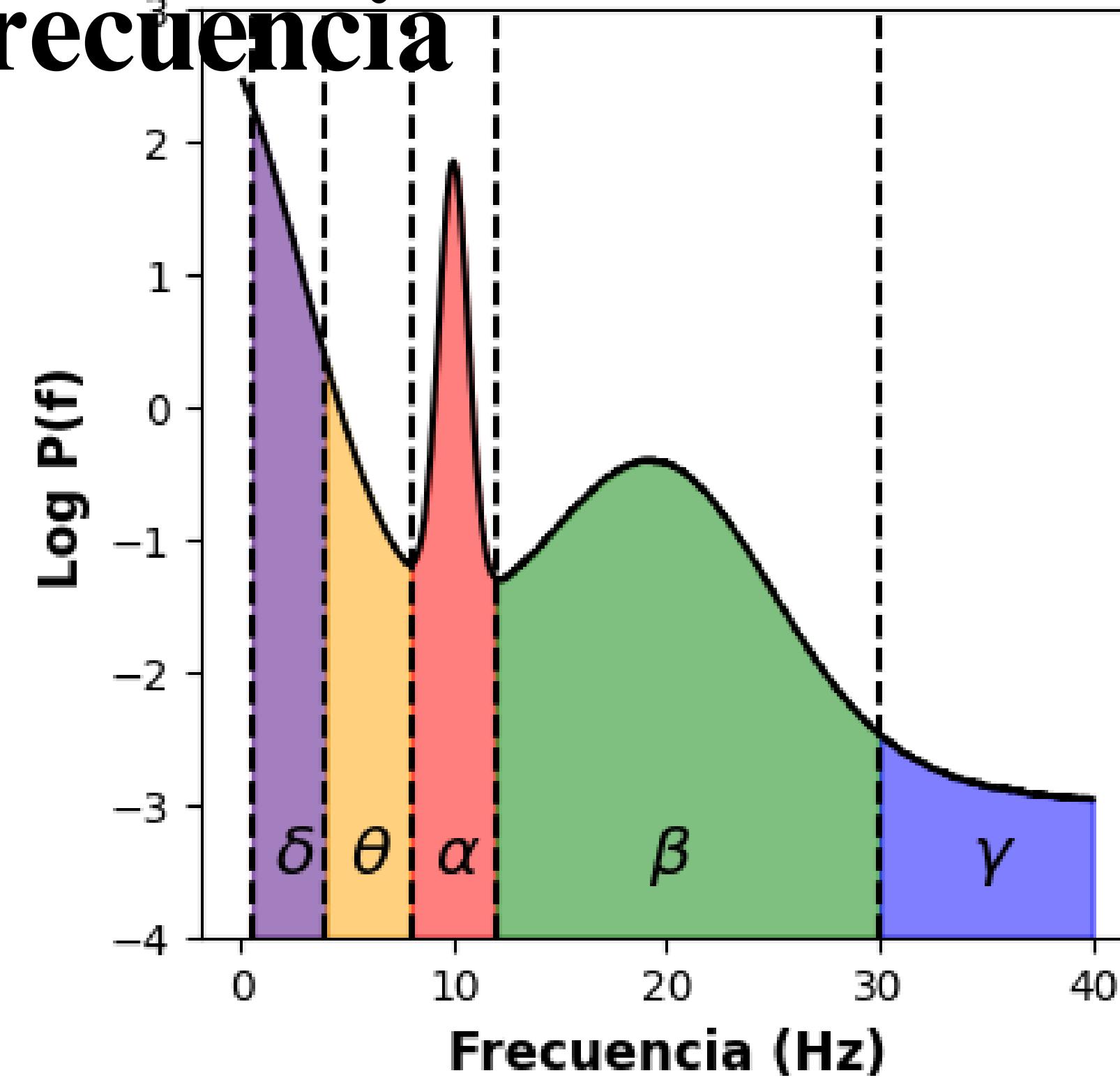
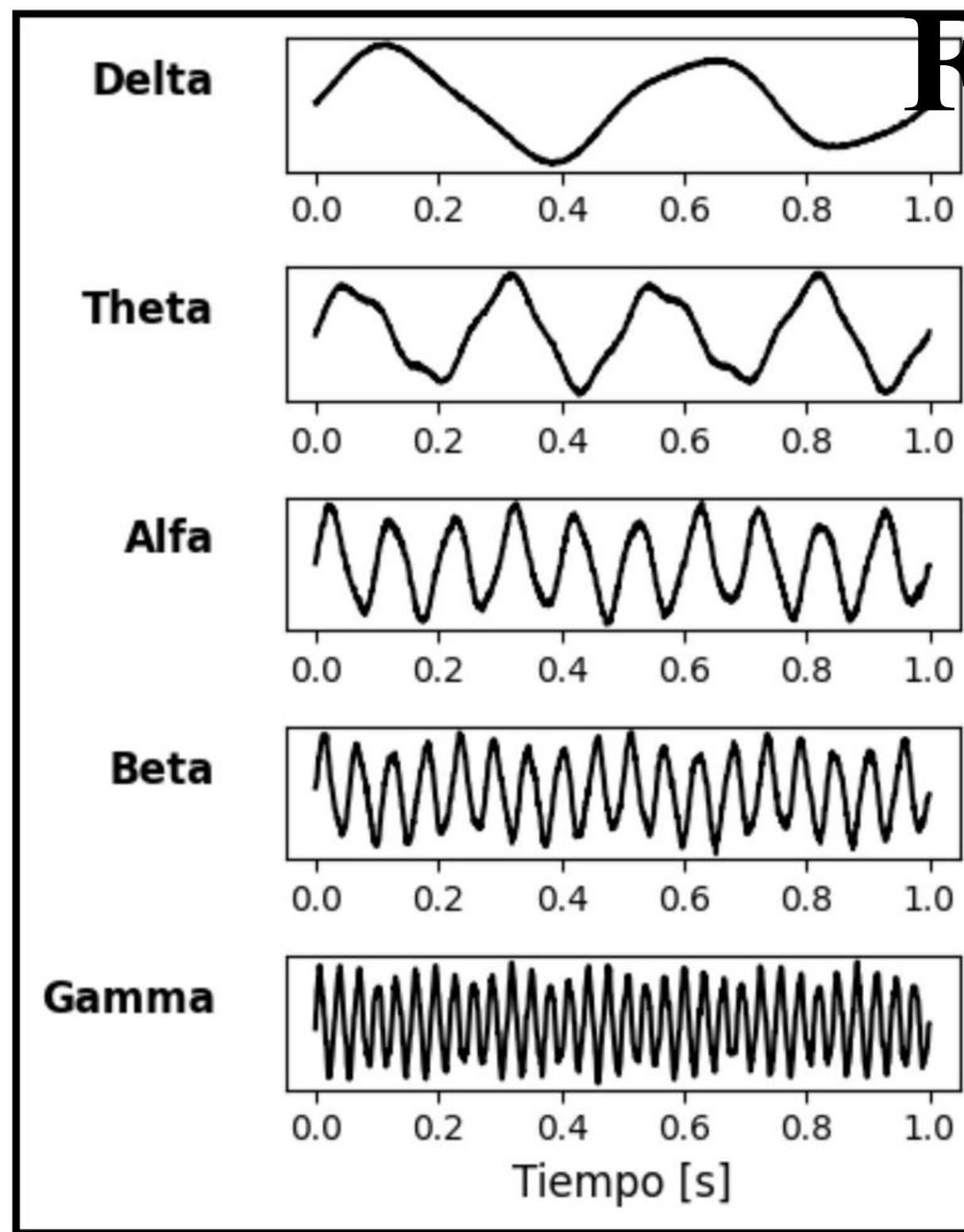


Electroencefalograma

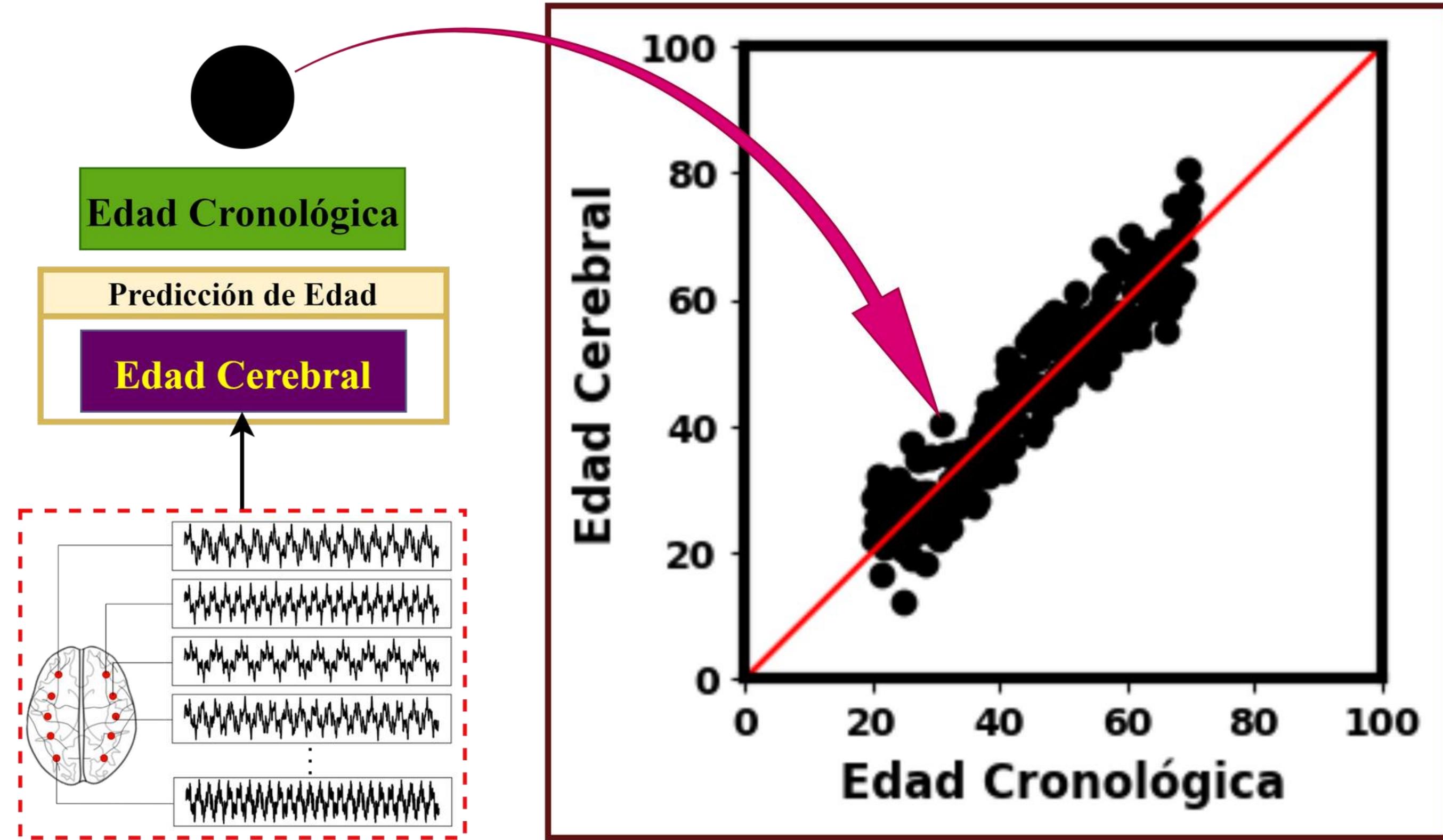
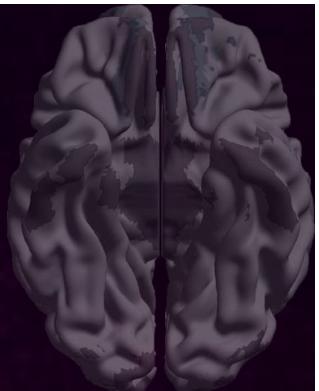


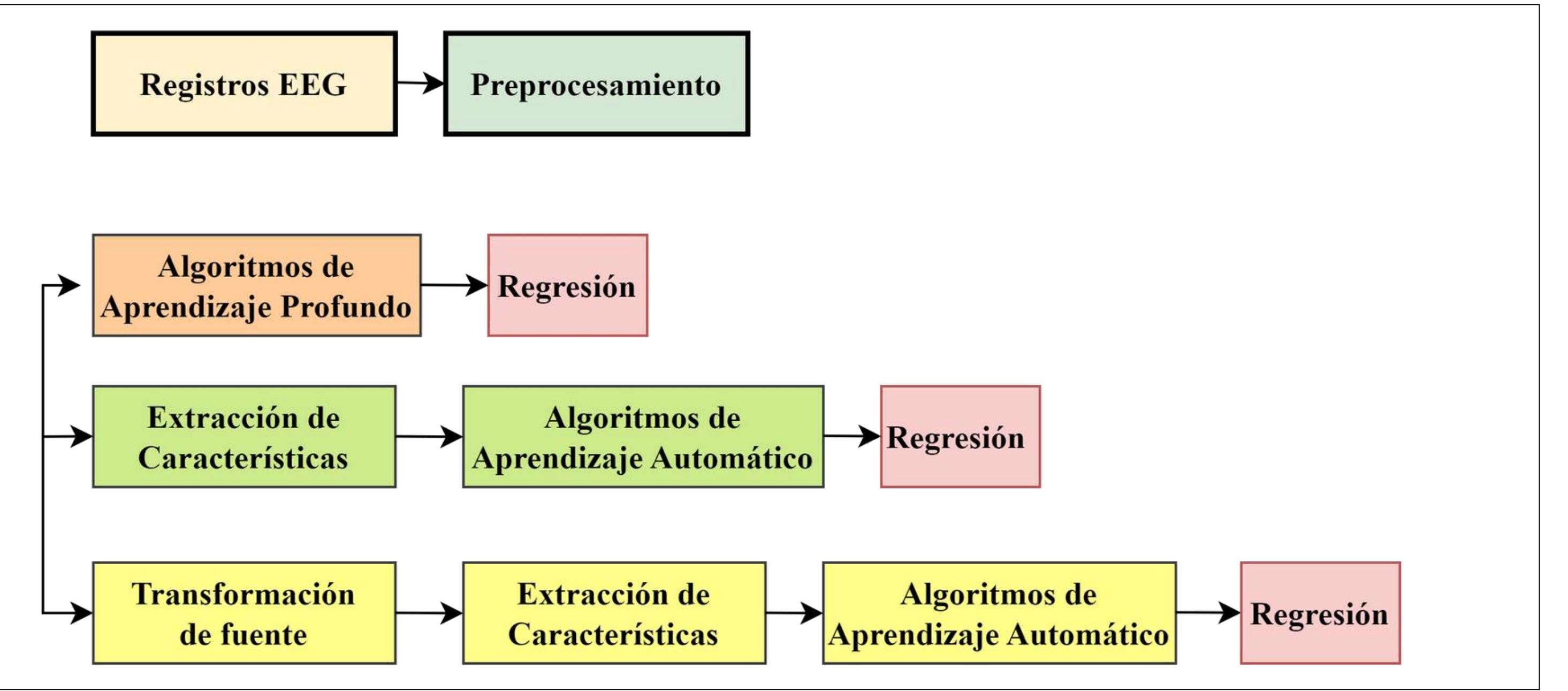
Espectro de

Frecuencia



Estimación de la Edad Cerebral

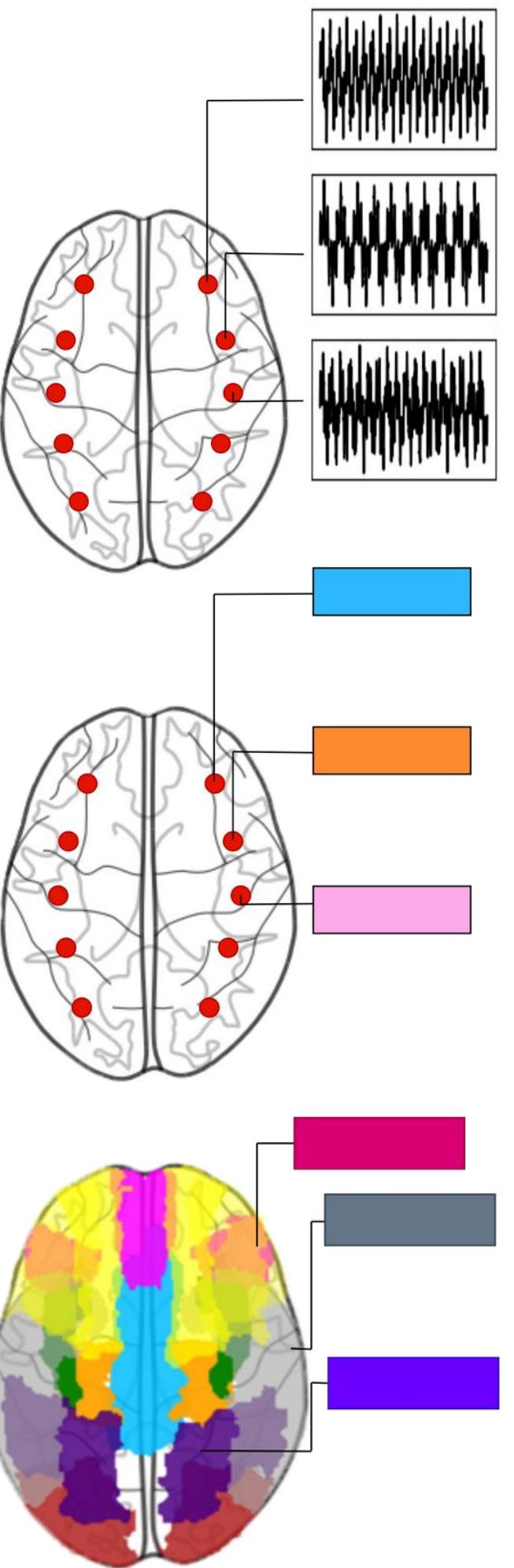


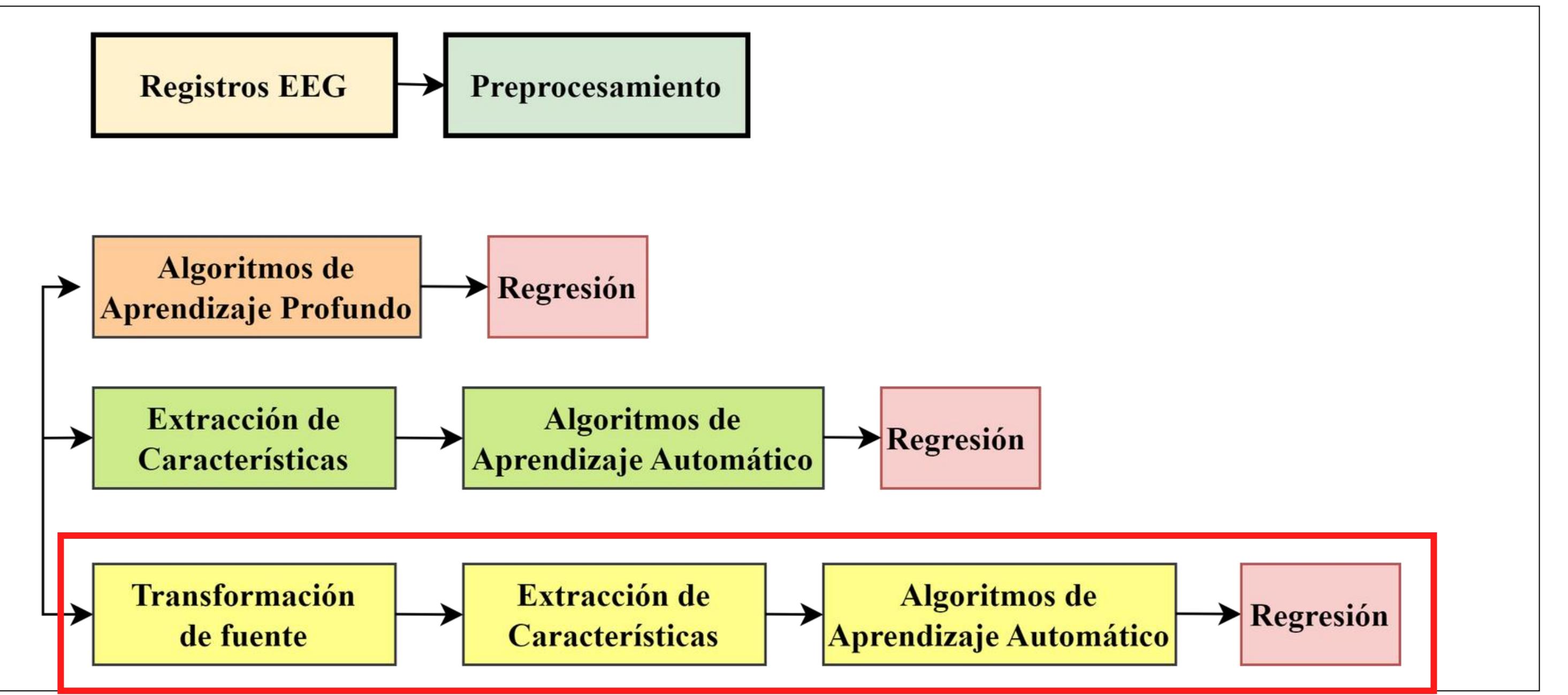


Khayretdinova, M. et al. Predicting age from resting-state scalp EEG signals with deep convolutional neural networks on TD-brain dataset. (2022).

Denis A. Engemann et al. “A reusable benchmark of brain-age prediction from M/EEG restingstate signals”(2022).

Obada Al Zoubi et al. “Predicting age from brain EEG signals-a machine learning approach”. (2018).

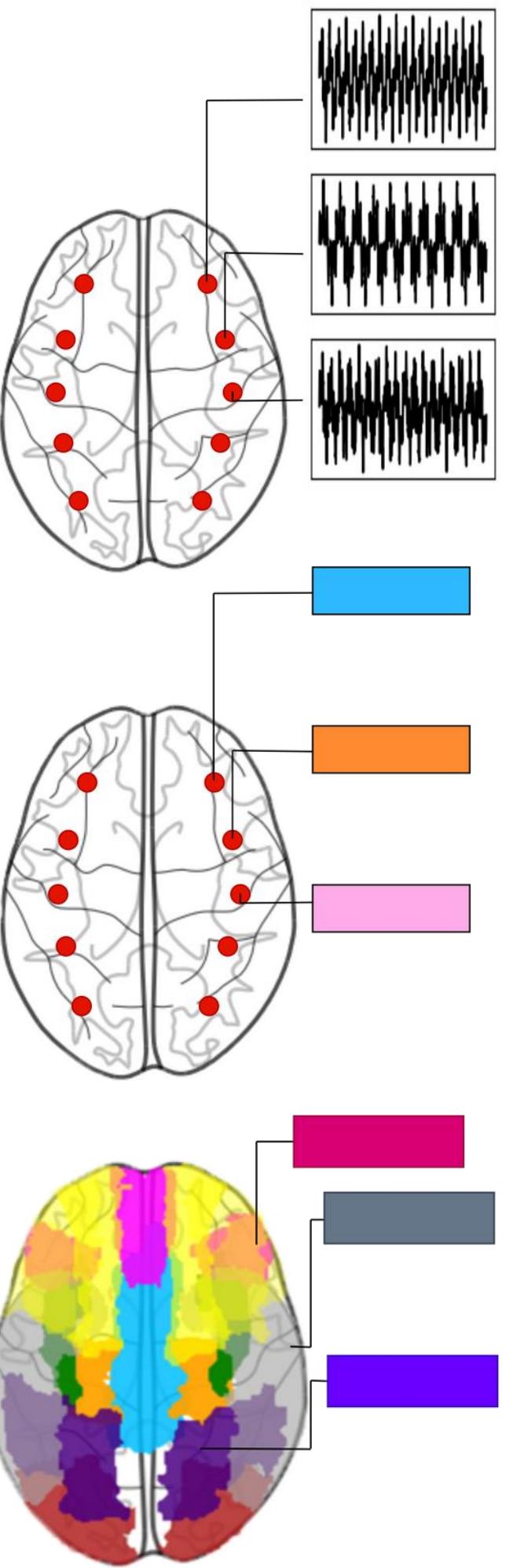




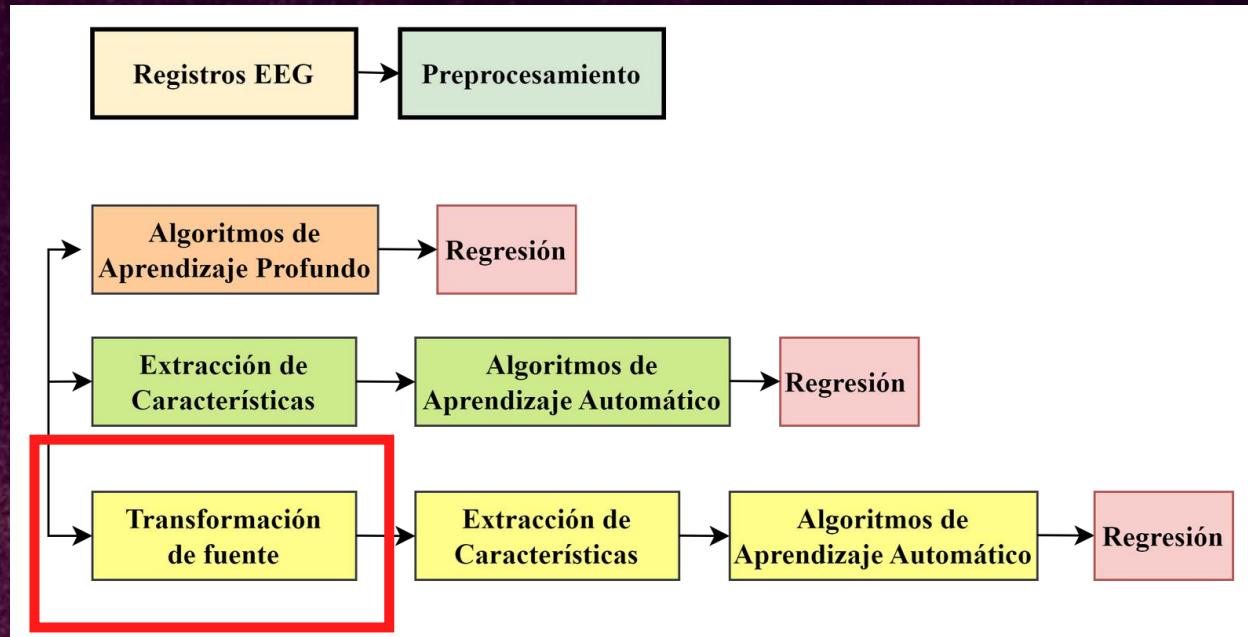
Khayretdinova, M. et al. Predicting age from resting-state scalp EEG signals with deep convolutional neural networks on TD-brain dataset. (2022).

Denis A. Engemann et al. “A reusable benchmark of brain-age prediction from M/EEG restingstate signals”(2022).

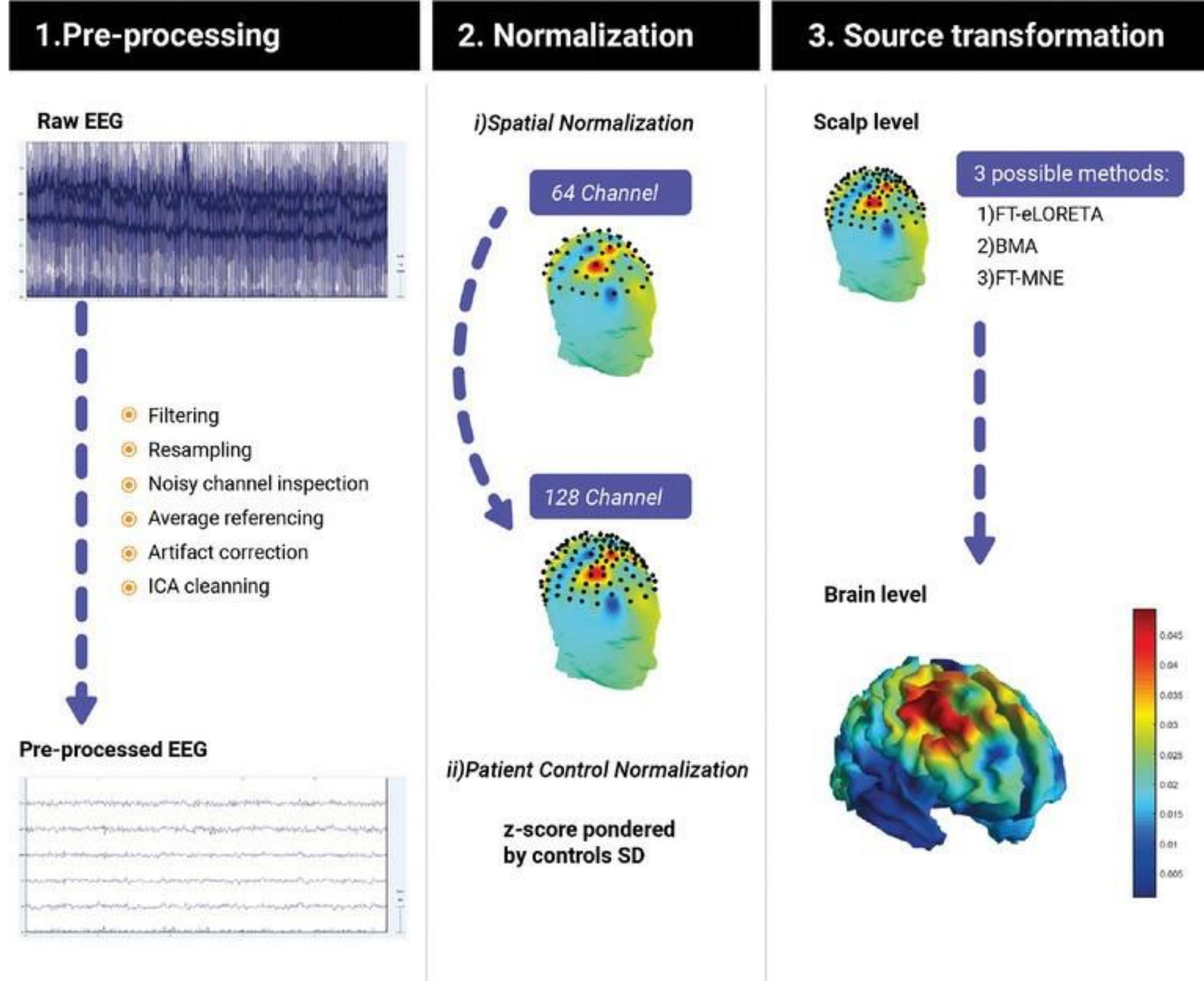
Obada Al Zoubi et al. “Predicting age from brain EEG signals-a machine learning approach”. (2018).



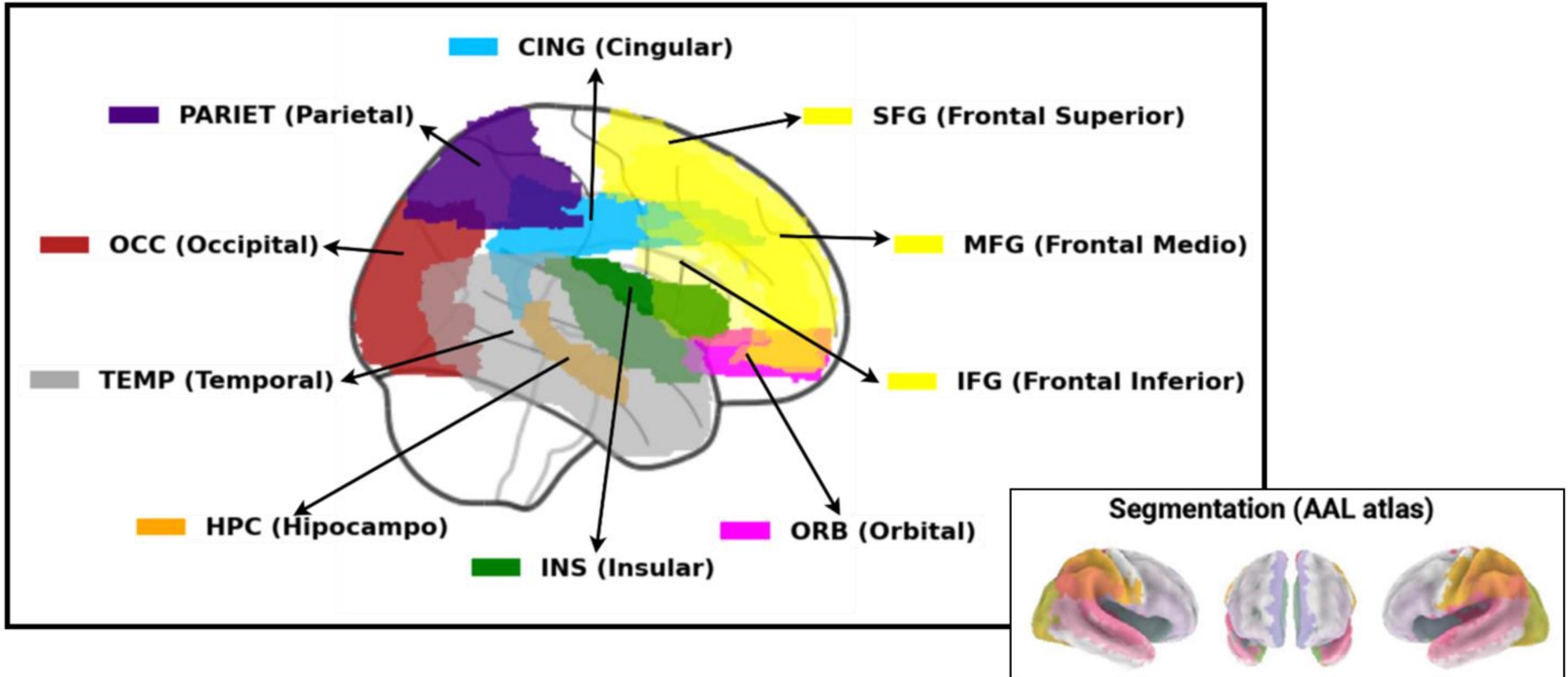
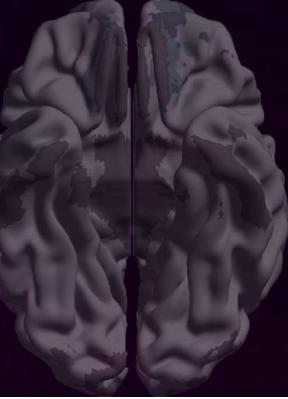
Transformación de Fuente



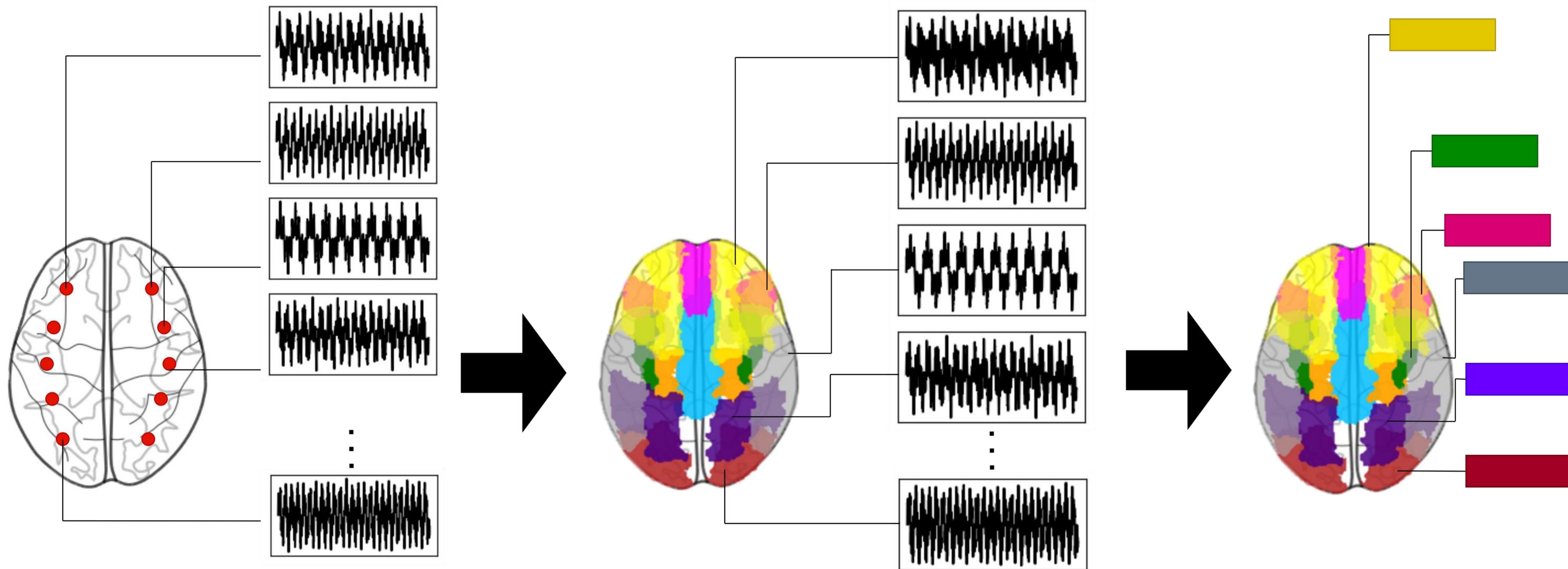
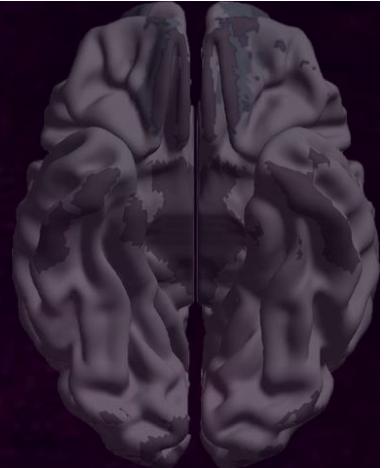
Prado, P. et al. Harmonized multi-metric and multi-centric assessment of EEG source space connectivity for dementia characterization. Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring 15. issn: 23528729 (3 jul. de 2023).



Modelo basado en regiones cerebrales



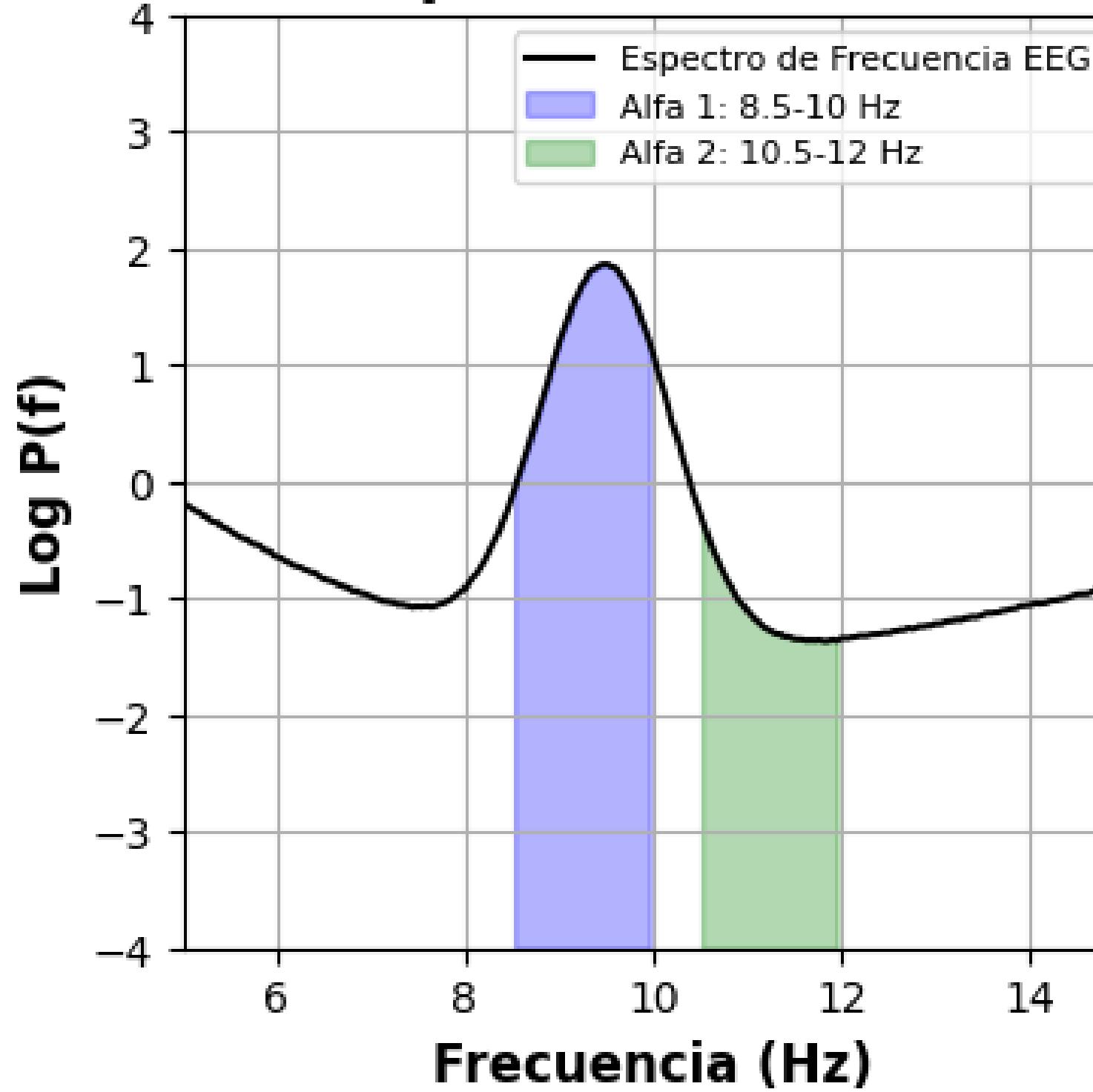
Modelo basado en regiones cerebrales



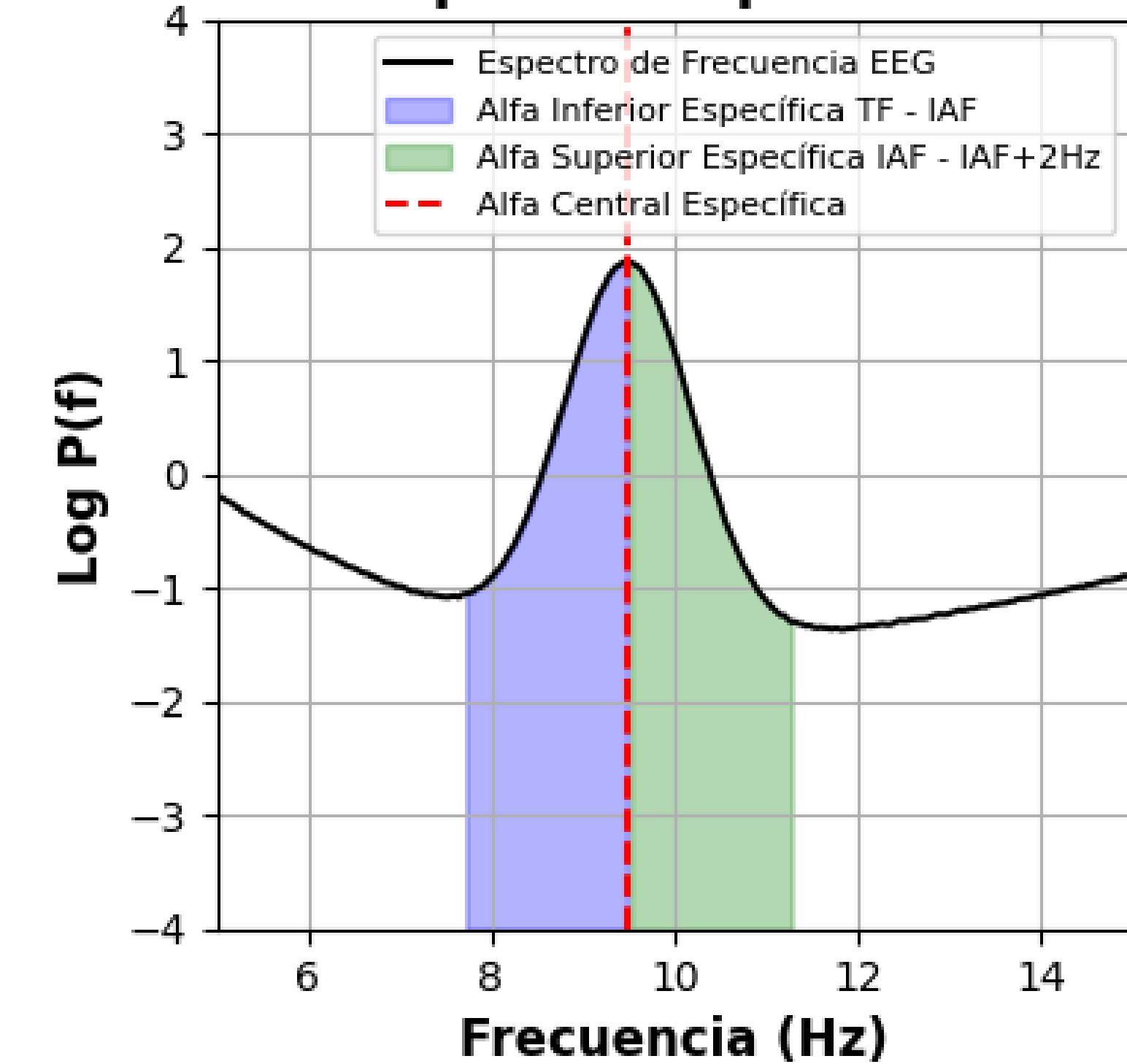
Espectro Canónico y Específico



Espectro Canónico



Espectro Específico



Extracción de Características

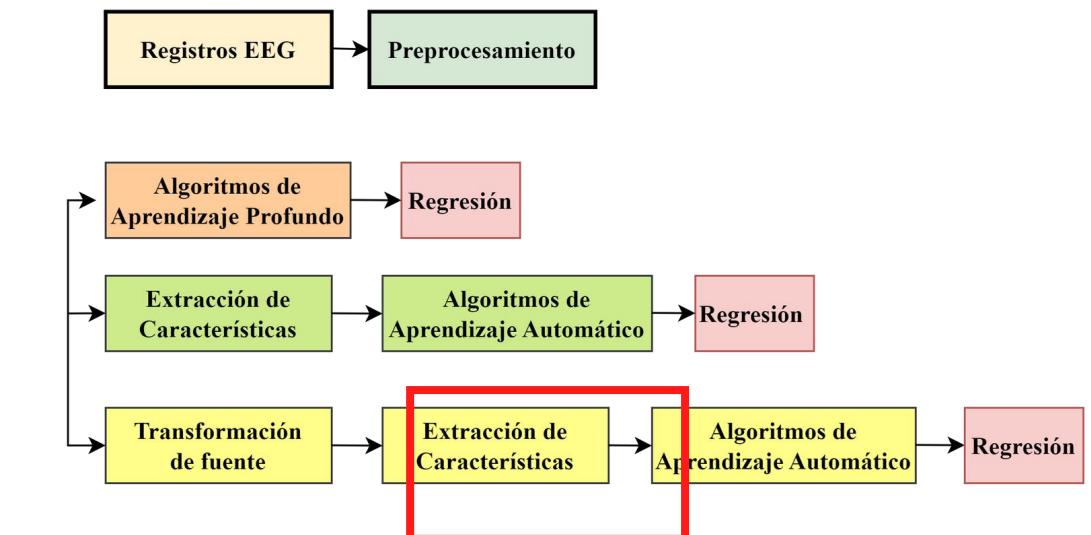
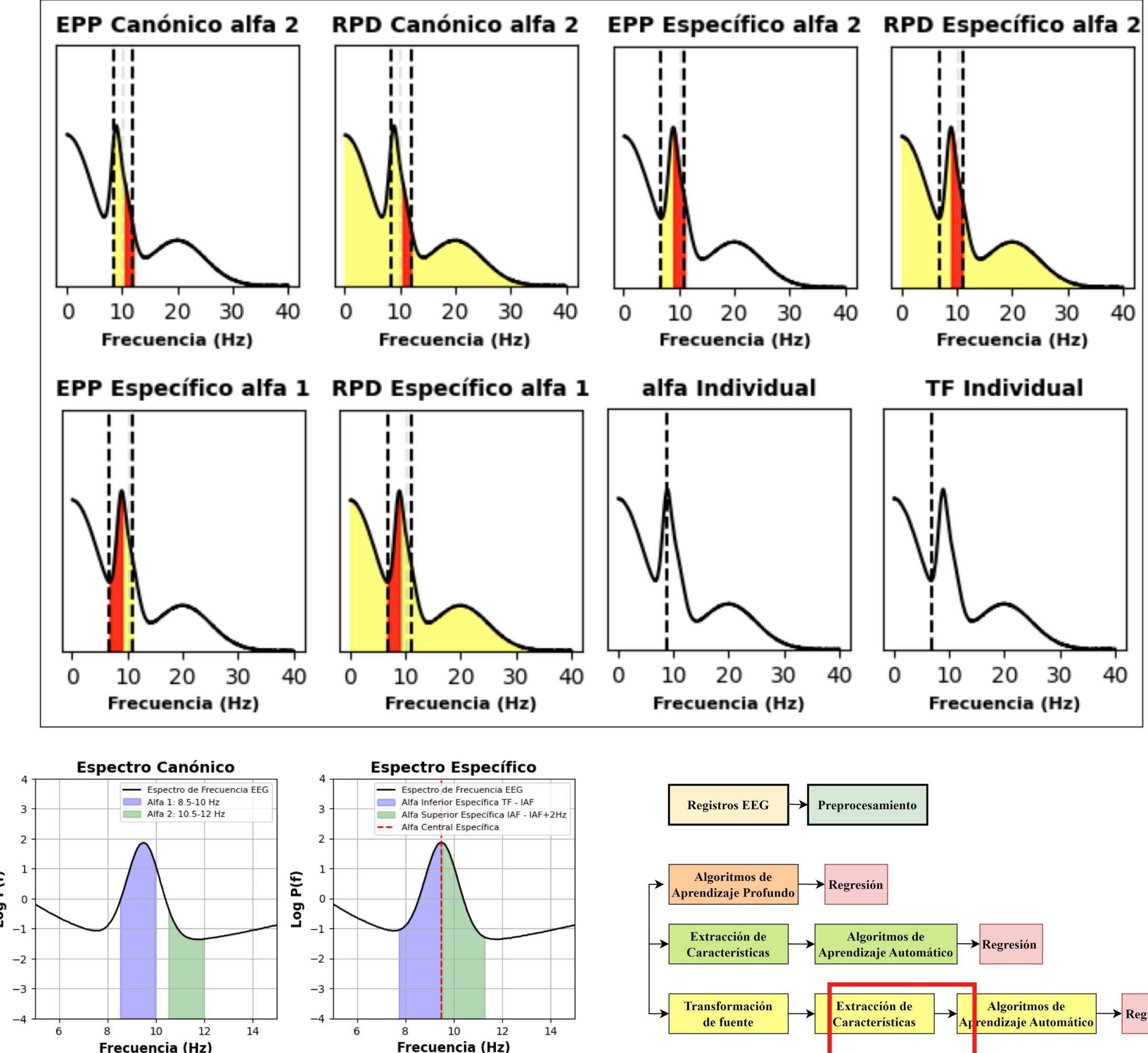
Densidad de Potencia Espectral (PSD):

$$\text{PSD}(f) = \frac{1}{T} |X(f)|^2$$

Densidad de potencia espectral normalizada (nPSD):

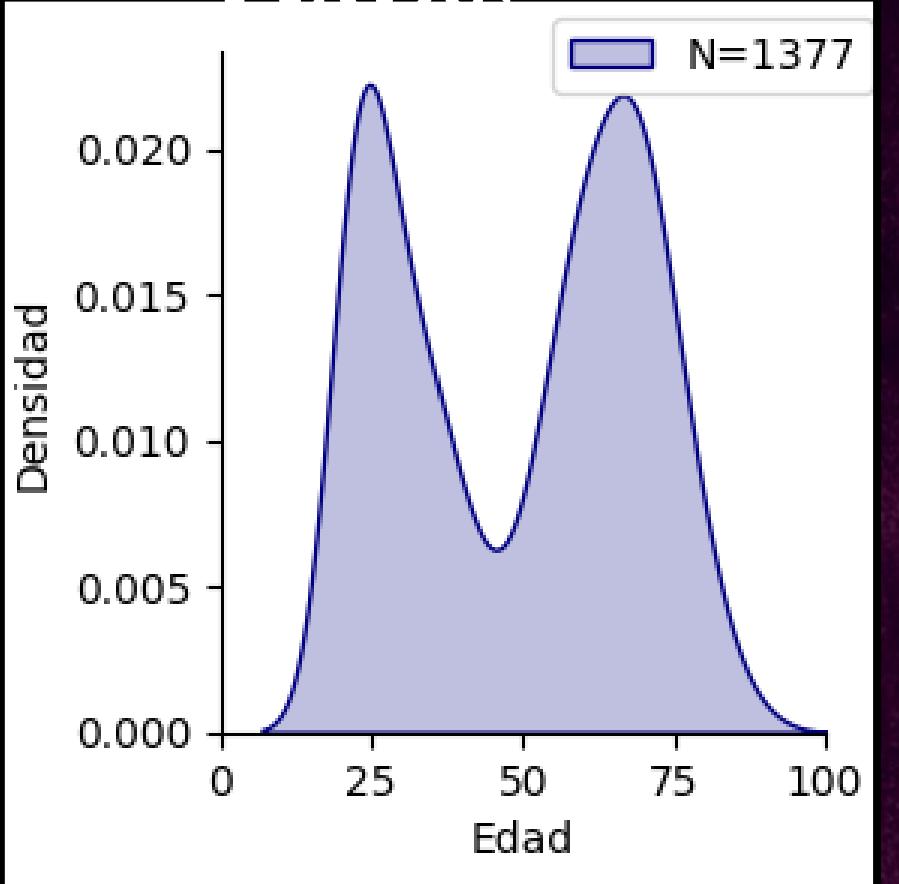
$$\text{nPSD}(f) = \frac{\text{PSD}(f)}{\sum_f \text{PSD}(f)}$$

EPP (Equivalent Percent Power)
RPD (Relative Power Density)



Base de

Datos



Distribución de edades de participantes del estudio.

Prado, P. et al. The BrainLat project, a multimodal neuroimaging dataset of neurodegeneration from underrepresented backgrounds. *Scientific Data* 10, 889. issn: 2052-4463.
<https://www.nature.com/articles/s41597-023-02806-8> (1 dic. de 2023).

Correlaciones (1)

Alpha2_canon_EPP_HPC_left	-0.13
Alpha2_canon_EPP_HPC_right	-0.12
Alpha2_canon_EPP_PARIET_left	-0.27
Alpha2_canon_EPP_PARIET_right	-0.27
Alpha2_canon_EPP_OCC_left	-0.19
Alpha2_canon_EPP_OCC_right	-0.22
Alpha2_canon_RPD_HPC_left	-0.13
Alpha2_canon_RPD_HPC_right	-0.12
Alpha2_canon_RPD_CING_left	-0.24
Alpha2_canon_RPD_CING_right	-0.27
Alpha2_canon_RPD_PARIET_left	-0.27
Alpha2_canon_RPD_PARIET_right	-0.27
Alpha2_canon_RPD_OCC_left	-0.19
Alpha2_canon_RPD_OCC_right	-0.22
High_subj_spec_EPP_HPC_left	-0.12
High_subj_spec_EPP_HPC_right	-0.11
High_subj_spec_EPP_CING_left	-0.24
High_subj_spec_EPP_CING_right	-0.27
High_subj_spec_EPP_PARIET_left	-0.26
High_subj_spec_EPP_PARIET_right	-0.26
High_subj_spec_EPP_OCC_left	-0.18
High_subj_spec_EPP_OCC_right	-0.20
High_subj_spec_RPD_HPC_left	-0.12
High_subj_spec_RPD_HPC_right	-0.11
High_subj_spec_RPD_CING_left	-0.24
High_subj_spec_RPD_CING_right	-0.27
High_subj_spec_RPD_PARIET_left	-0.26
High_subj_spec_RPD_PARIET_right	-0.26
High_subj_spec_RPD_OCC_left	-0.18
High_subj_spec_RPD_OCC_right	-0.20
Low_subj_spec_EPP_HPC_left	-0.26
Low_subj_spec_EPP_HPC_right	-0.22
Low_subj_spec_EPP_CING_left	-0.30
Low_subj_spec_EPP_CING_right	-0.32
Low_subj_spec_EPP_PARIET_left	-0.37

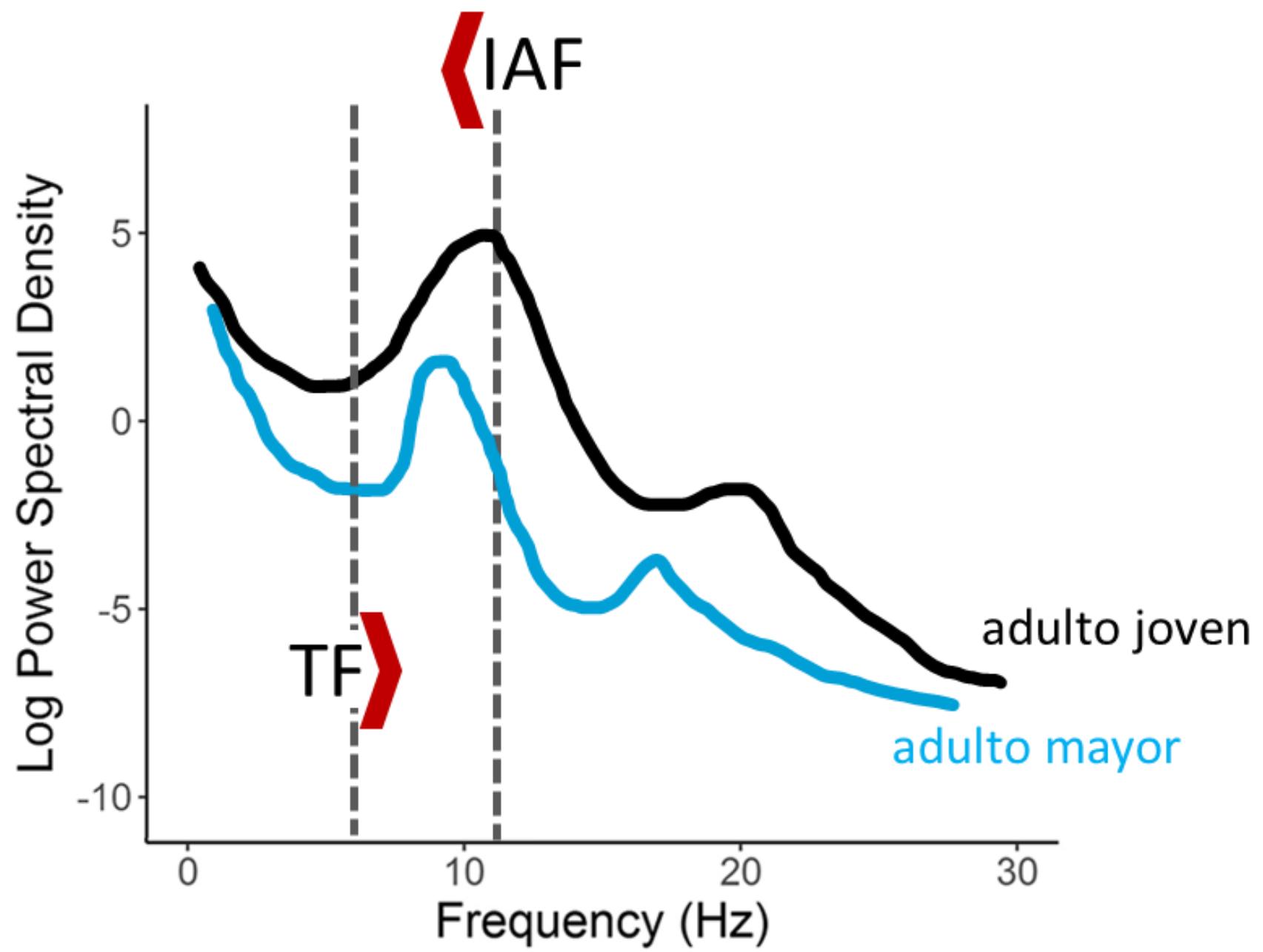
Edad

Correlaciones (2)

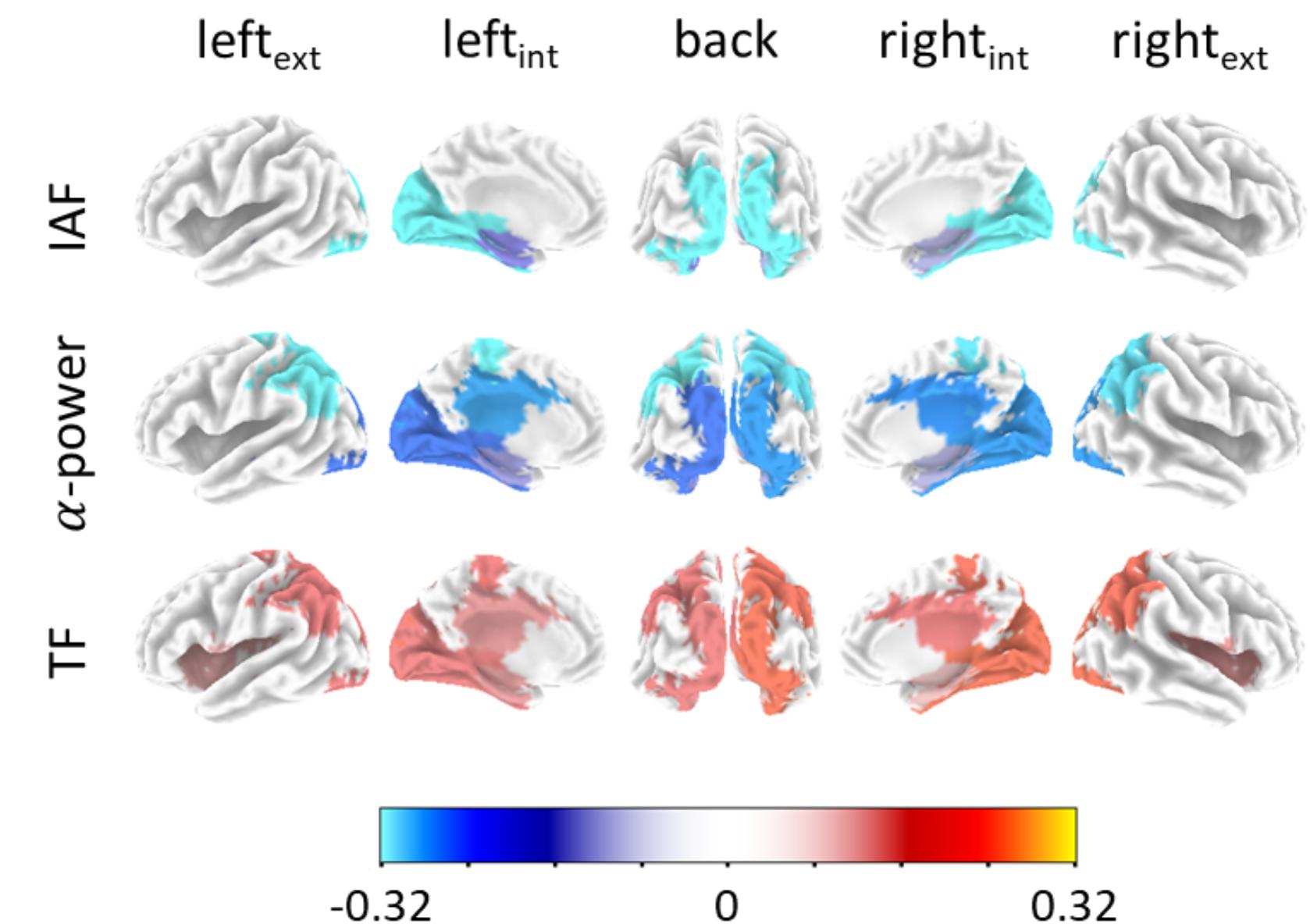
Low_subj_spec_EPP_OCC_left	-0.34
Low_subj_spec_EPP_OCC_right	-0.36
Low_subj_spec_RPD_HPC_left	-0.09
Low_subj_spec_RPD_HPC_right	-0.08
Low_subj_spec_RPD_CING_left	-0.22
Low_subj_spec_RPD_CING_right	-0.24
Low_subj_spec_RPD_PARIET_left	-0.24
Low_subj_spec_RPD_PARIET_right	-0.22
Low_subj_spec_RPD_OCC_left	-0.13
Low_subj_spec_RPD_OCC_right	-0.15
TF_ORB_left	0.21
TF_ORB_right	0.22
TF_IFG_left	0.19
TF_IFG_right	0.22
TF_MFG_left	0.22
TF_MFG_right	0.25
TF_SFG_left	0.22
TF_SFG_right	0.25
TF_HPC_left	0.19
TF_HPC_right	0.16
TF_INS_left	0.18
TF_INS_right	0.20
IAF_ORB_left	-0.15
IAF_ORB_right	-0.19
IAF_IFG_left	-0.16
IAF_IFG_right	-0.16
IAF_MFG_left	-0.18
IAF_MFG_right	-0.17
IAF_SFG_left	-0.20
IAF_SFG_right	-0.20
IAF_HPC_left	-0.21
IAF_HPC_right	-0.17
IAF_INS_left	-0.19
IAF_INS_right	-0.20
IAF_OCC_left	-0.34
IAF_OCC_right	-0.35

14

Edad

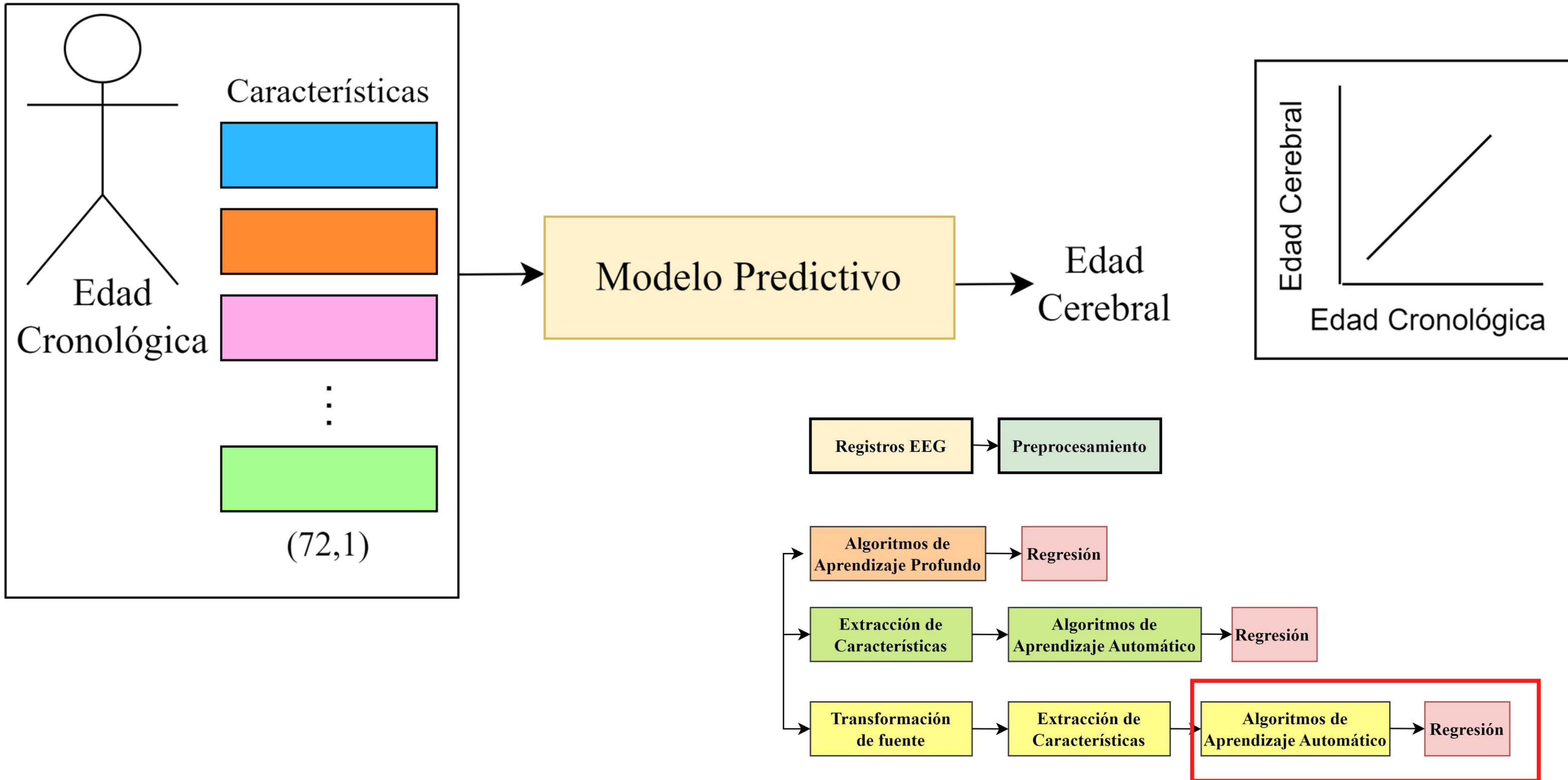


Cambios espectrales en el EEG
durante el envejecimiento



Correlación con la edad de diferentes
descriptores de las oscilaciones α

Modelo Predictivo



Métricas de Evaluación

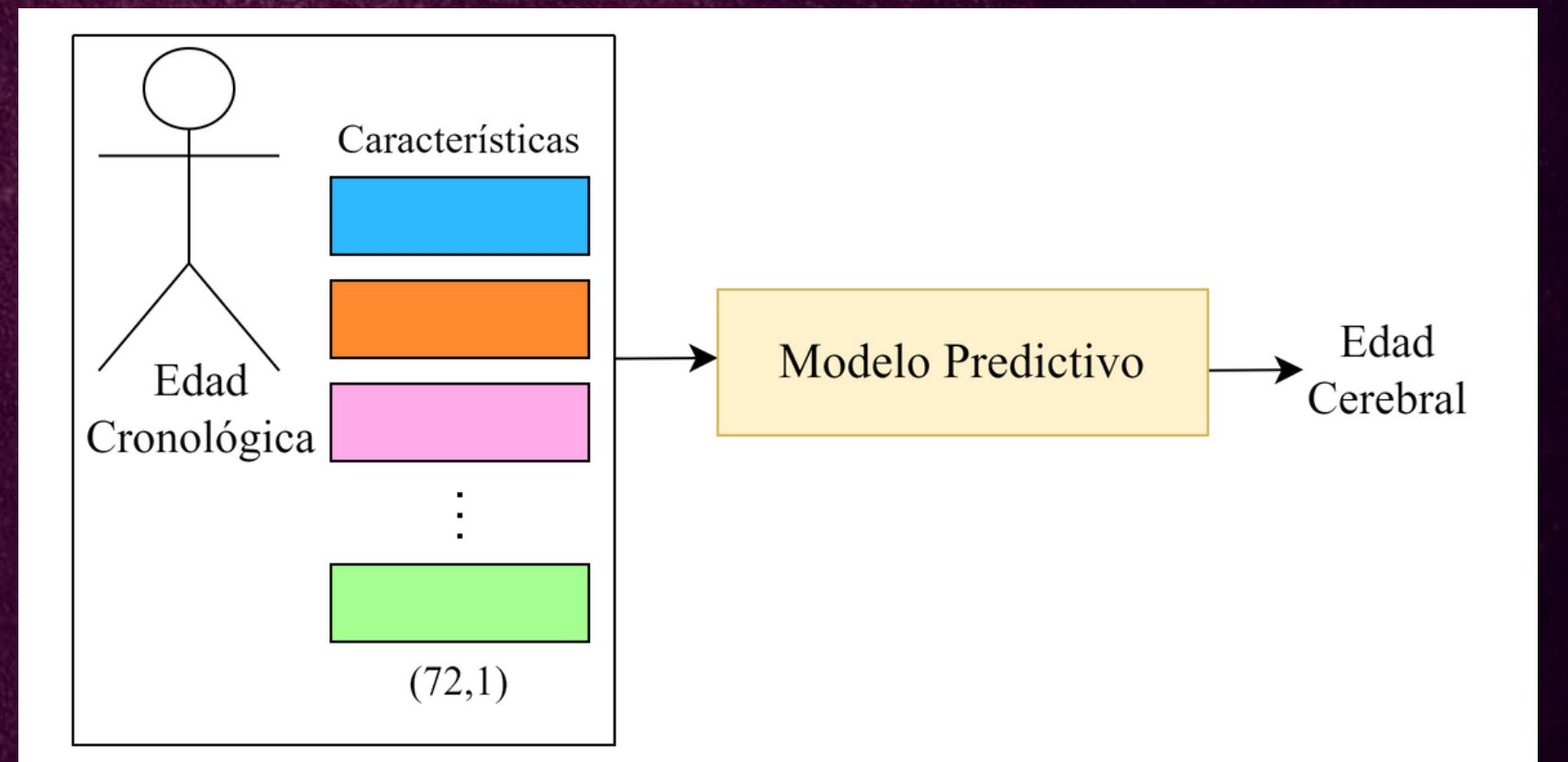
MSE: Error Cuadrático Medio

MAE: Error Absoluto Medio

R²: Coeficiente de Determinación

y: Edad Cronológica

\hat{y} : Edad Cerebral

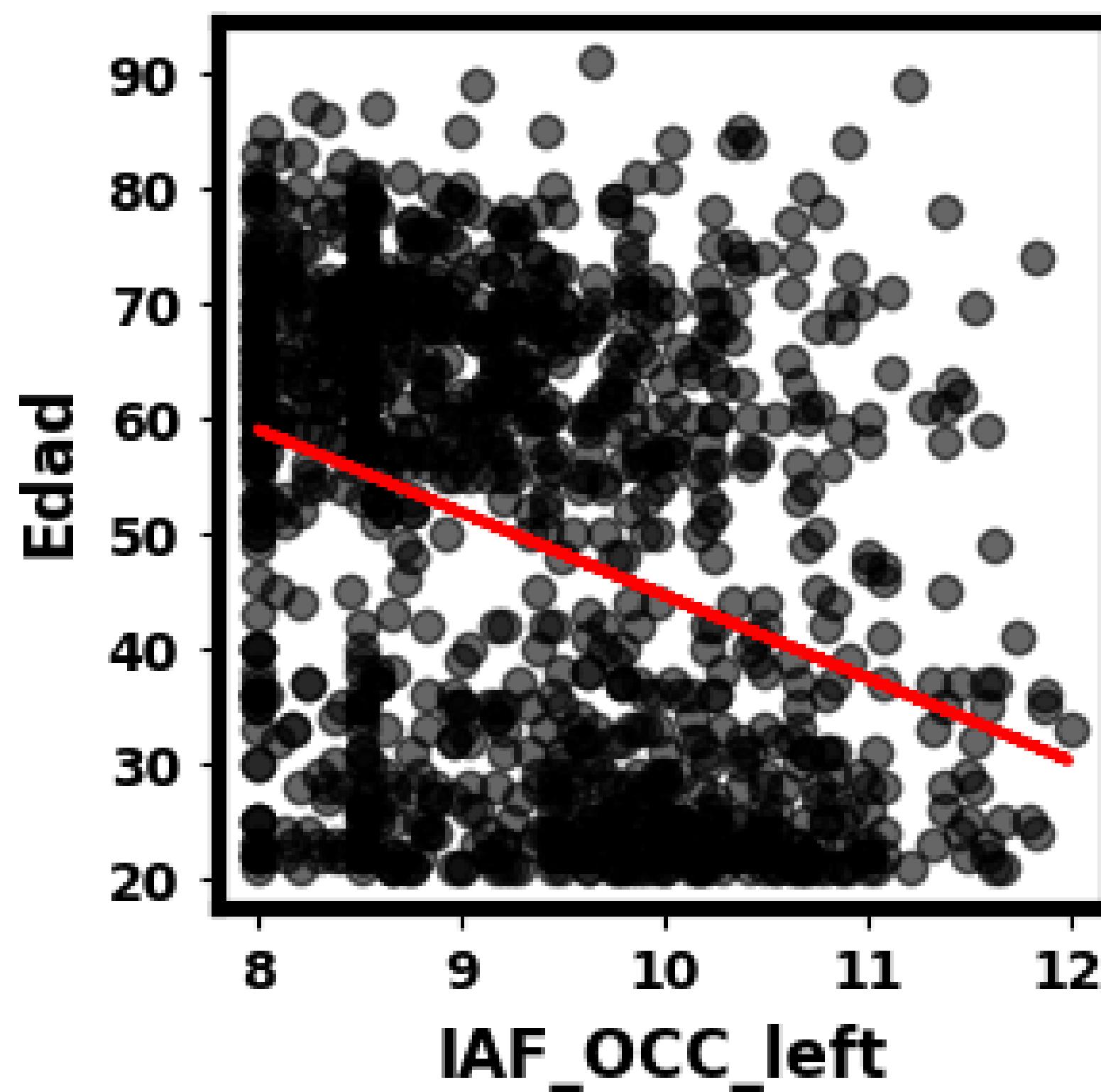


$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

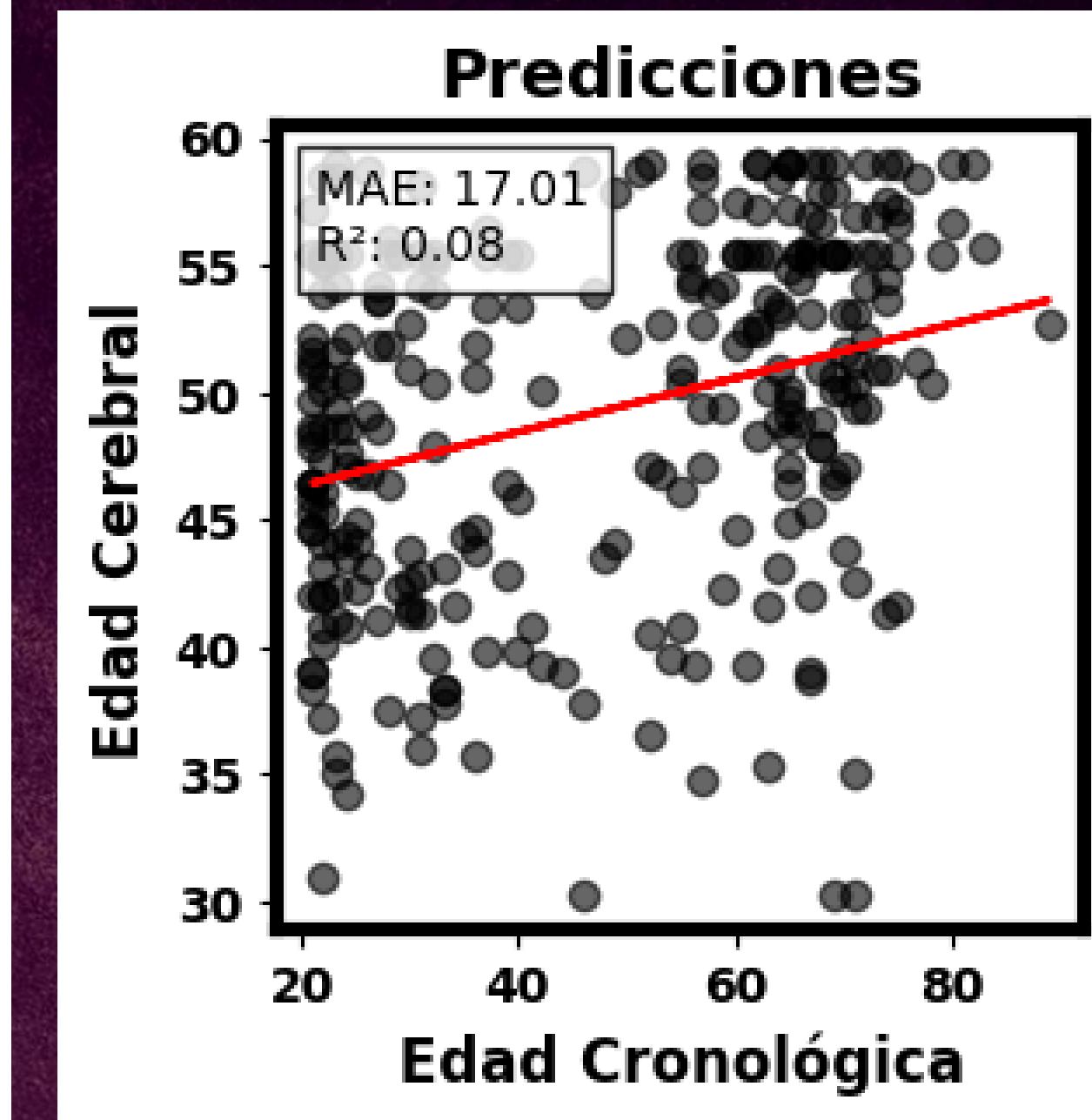
$$R^2 = 1 - \frac{MSE}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

Datos de Entrenamiento

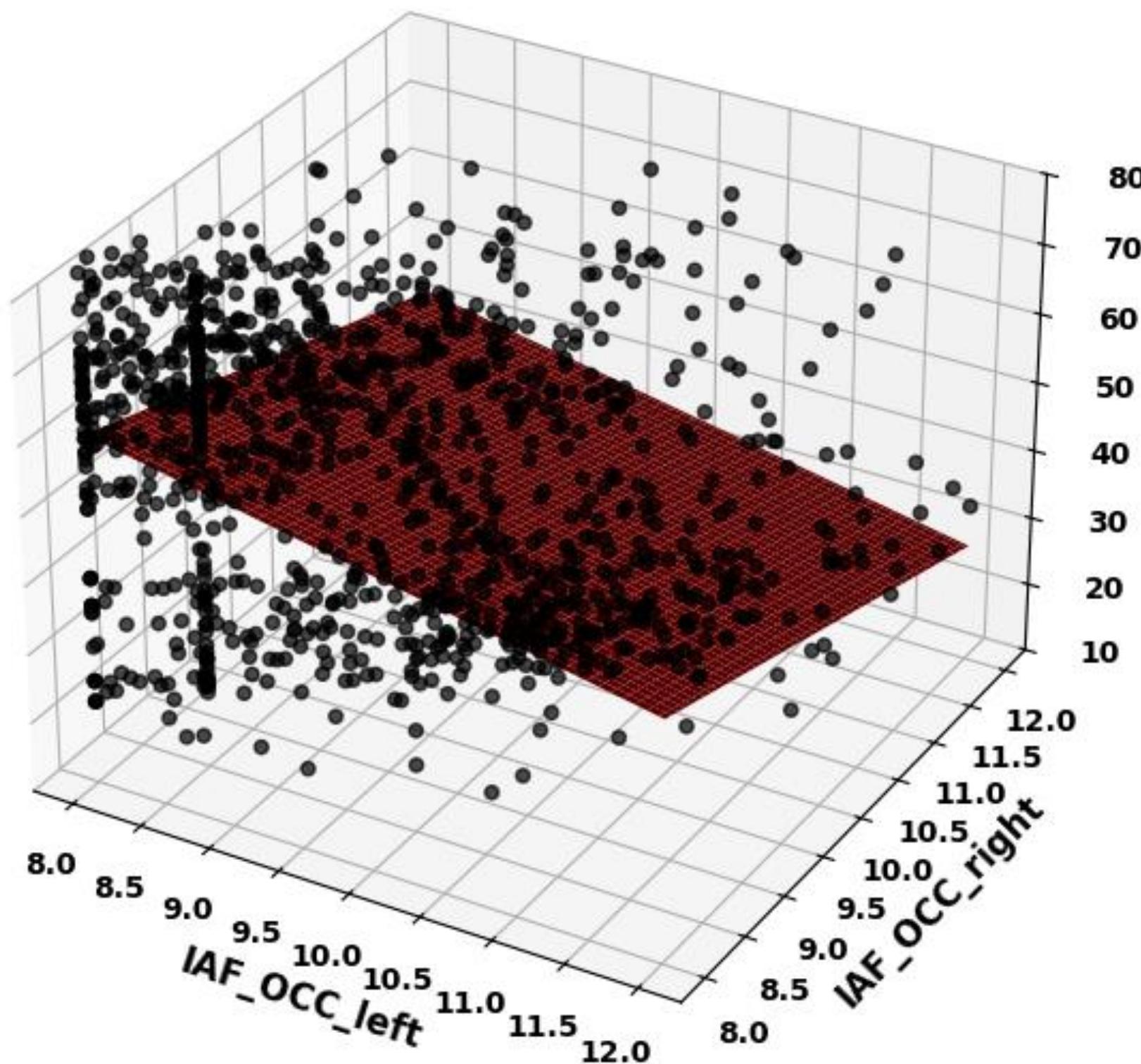


$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Modelo con 1 característica:

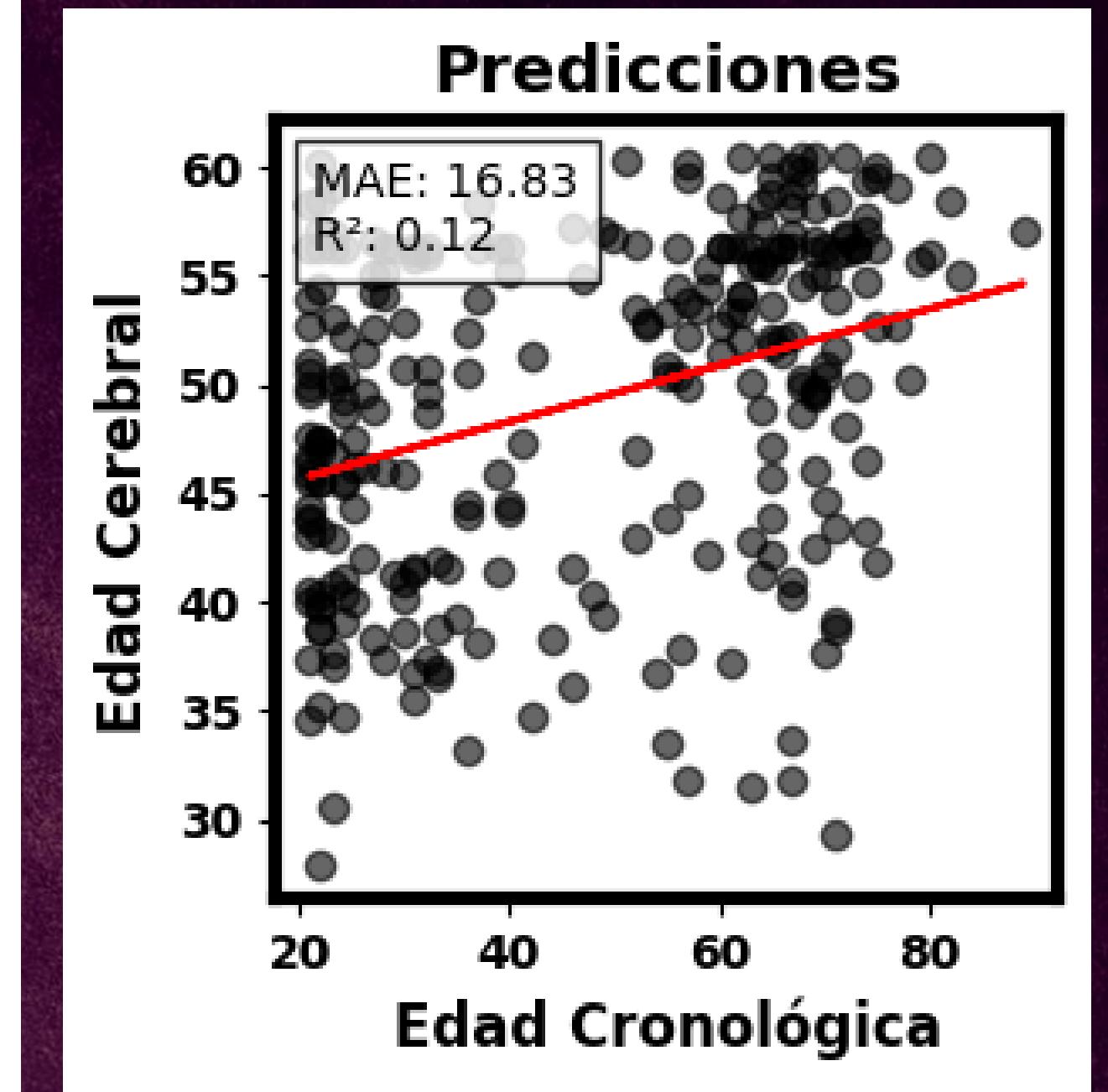


Modelo Ajustado a IAF_OCC_left y IAF_OCC_right



$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_m x_{i,m} + \epsilon_i$$

Modelo con 2 características:

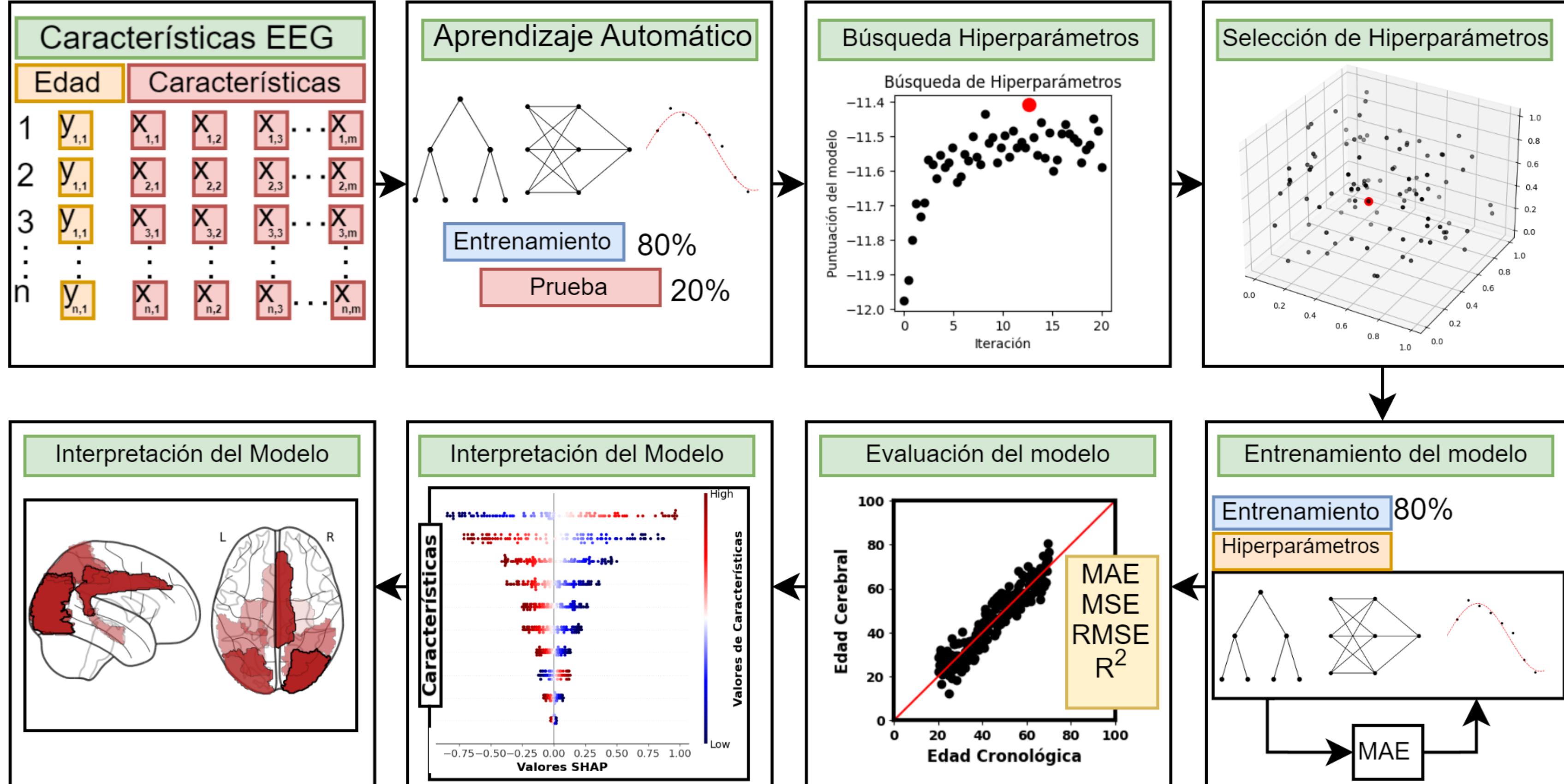
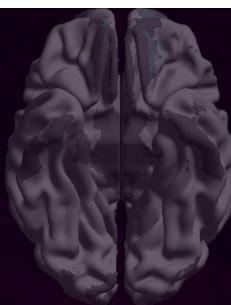


Modelo con 72 características:

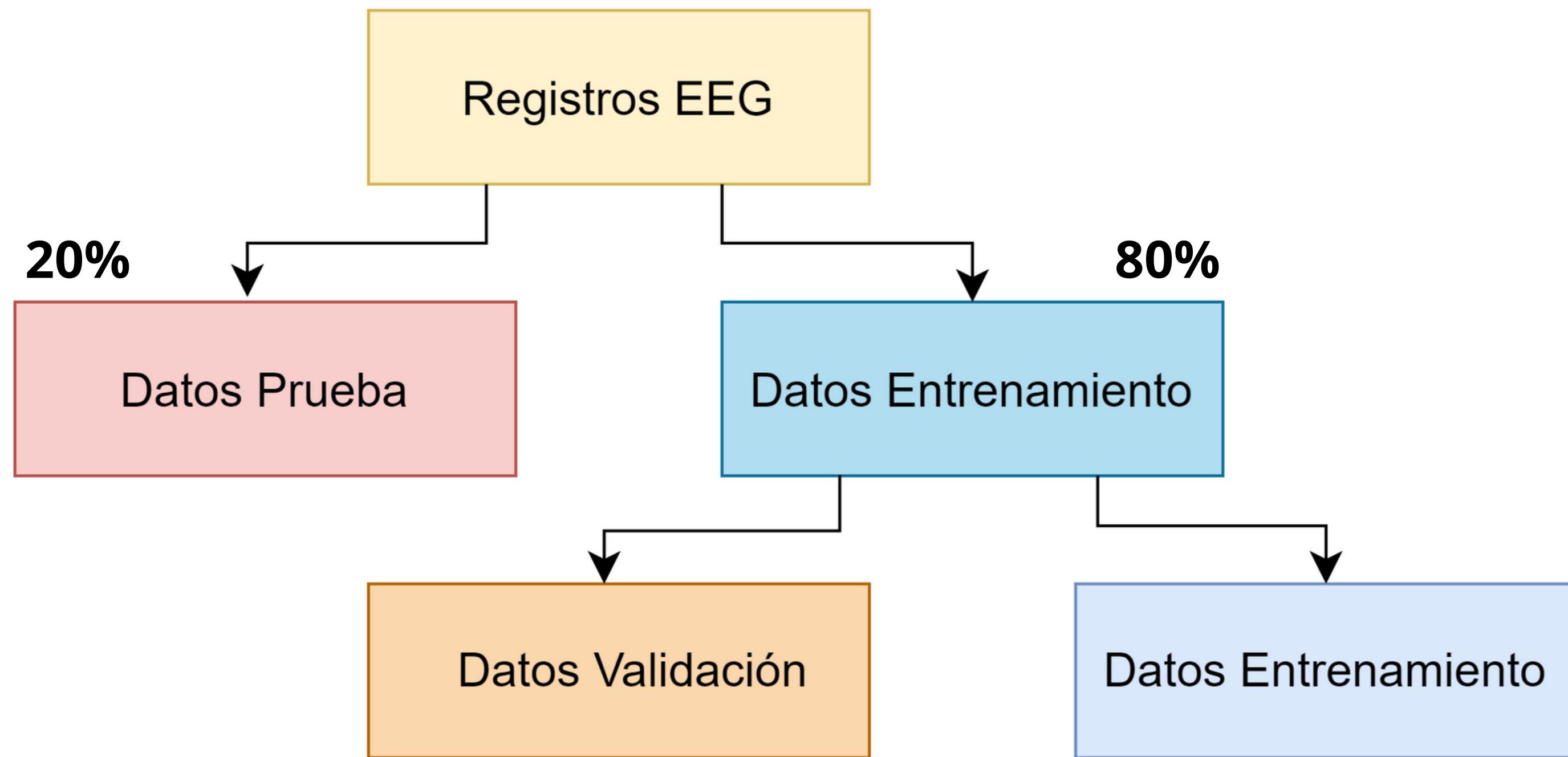
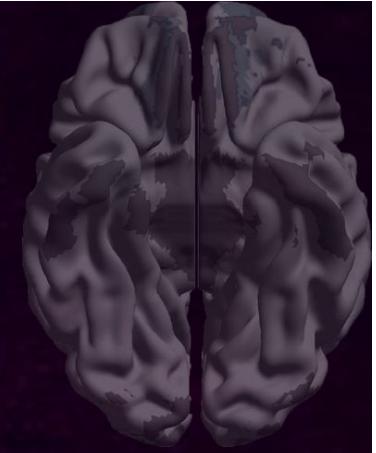
R²=0.24

MAE=14.57

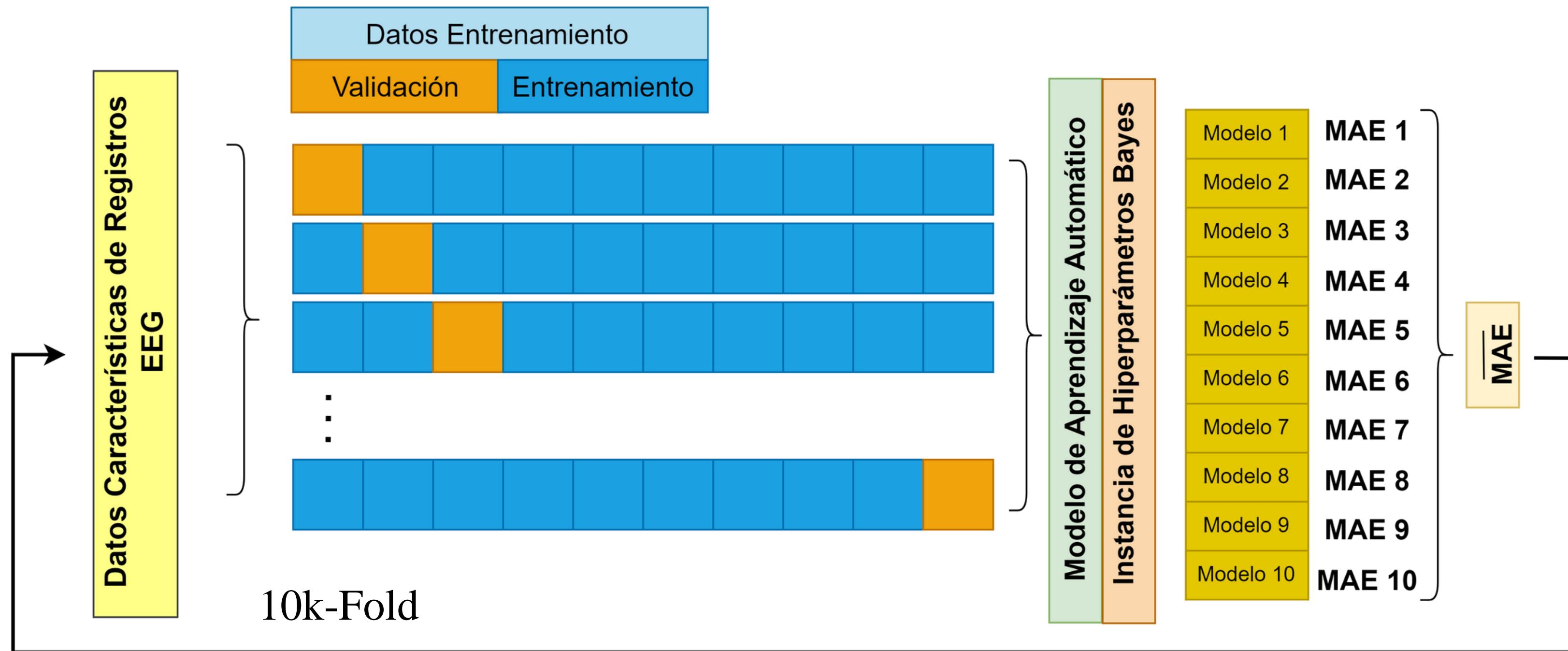
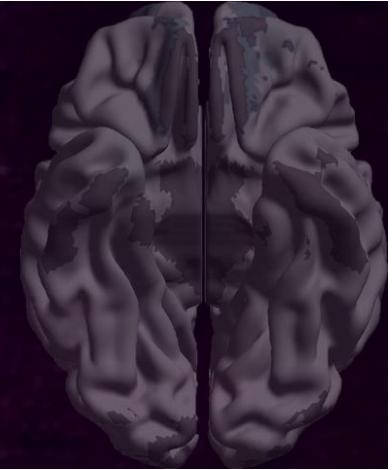
Modelo de Regresión



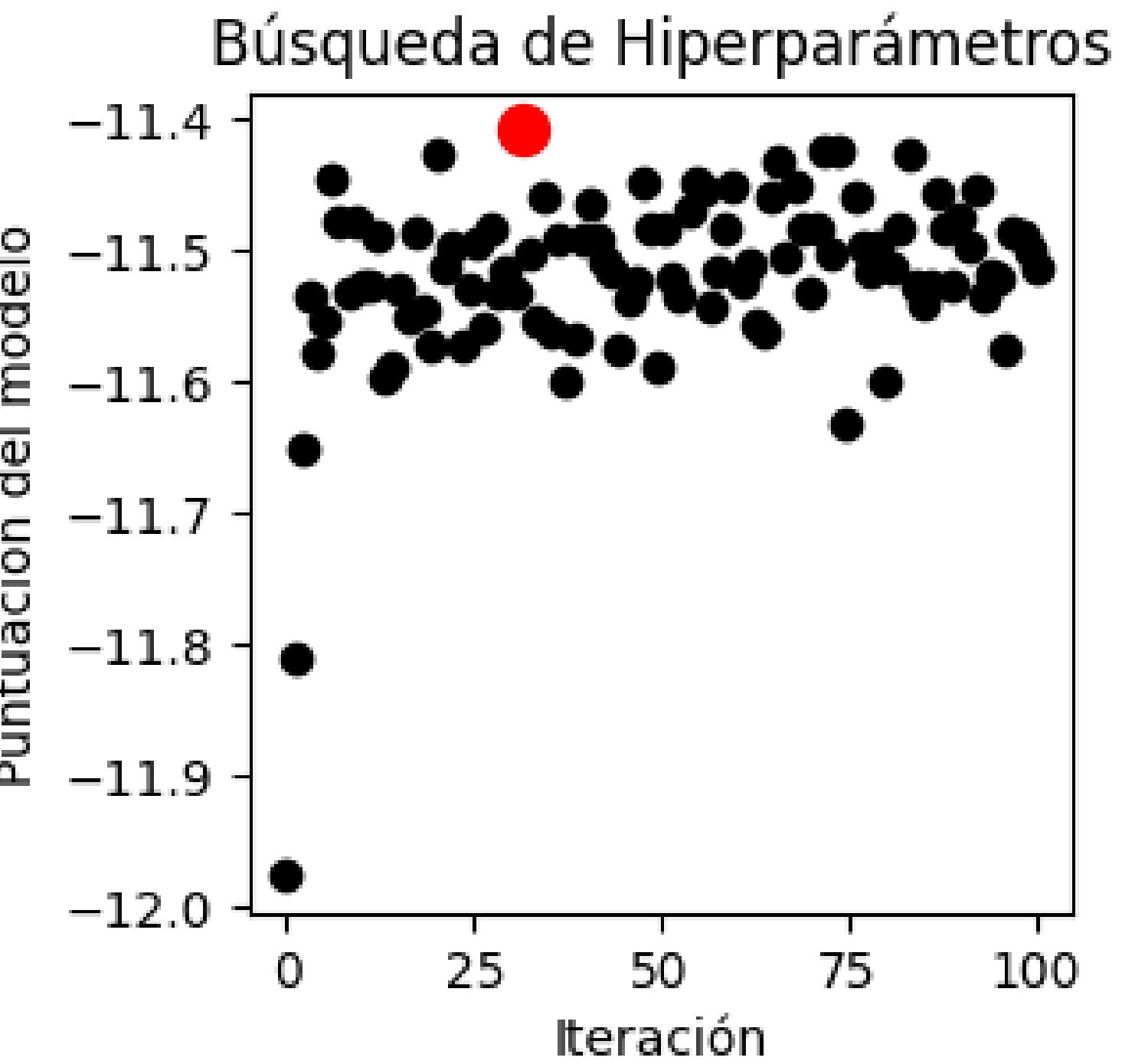
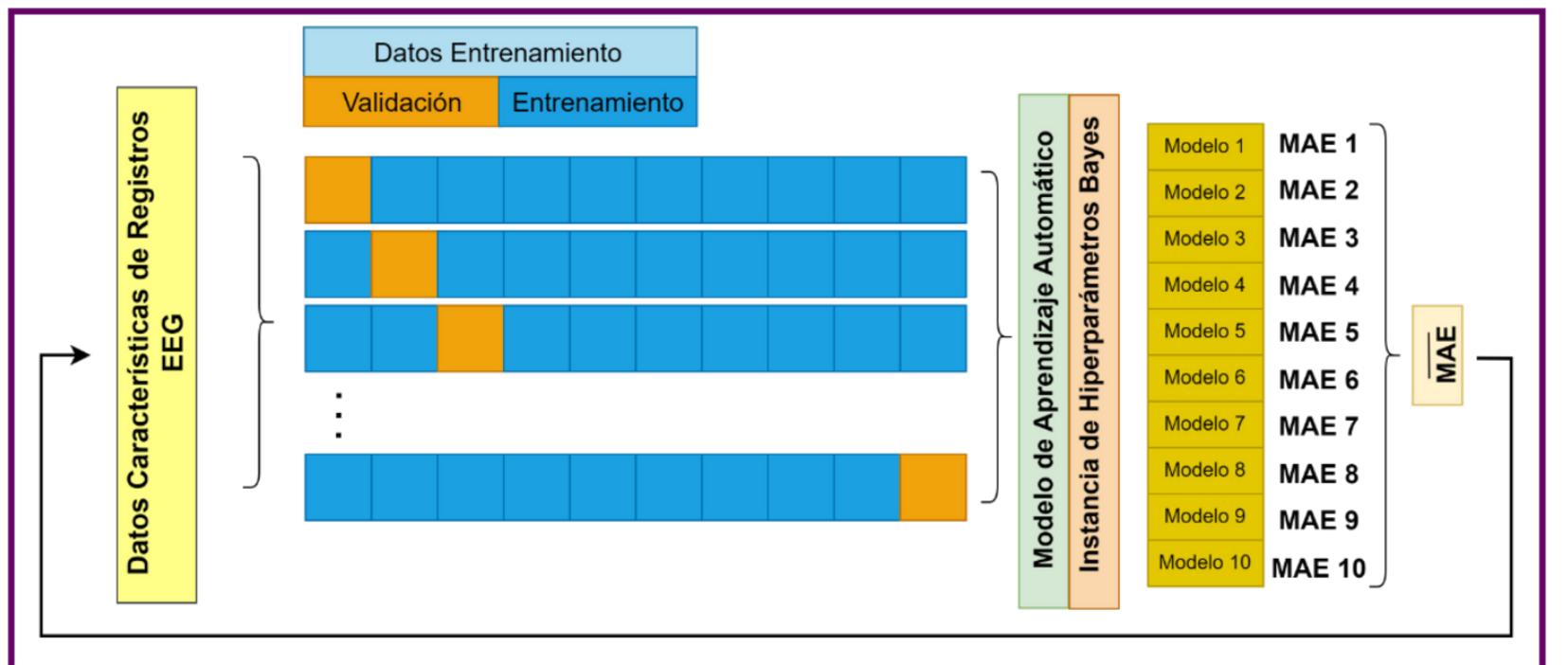
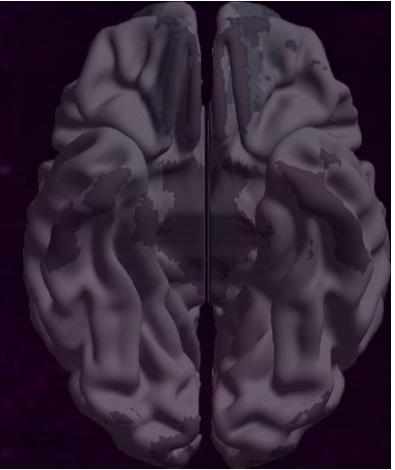
División de los Datos



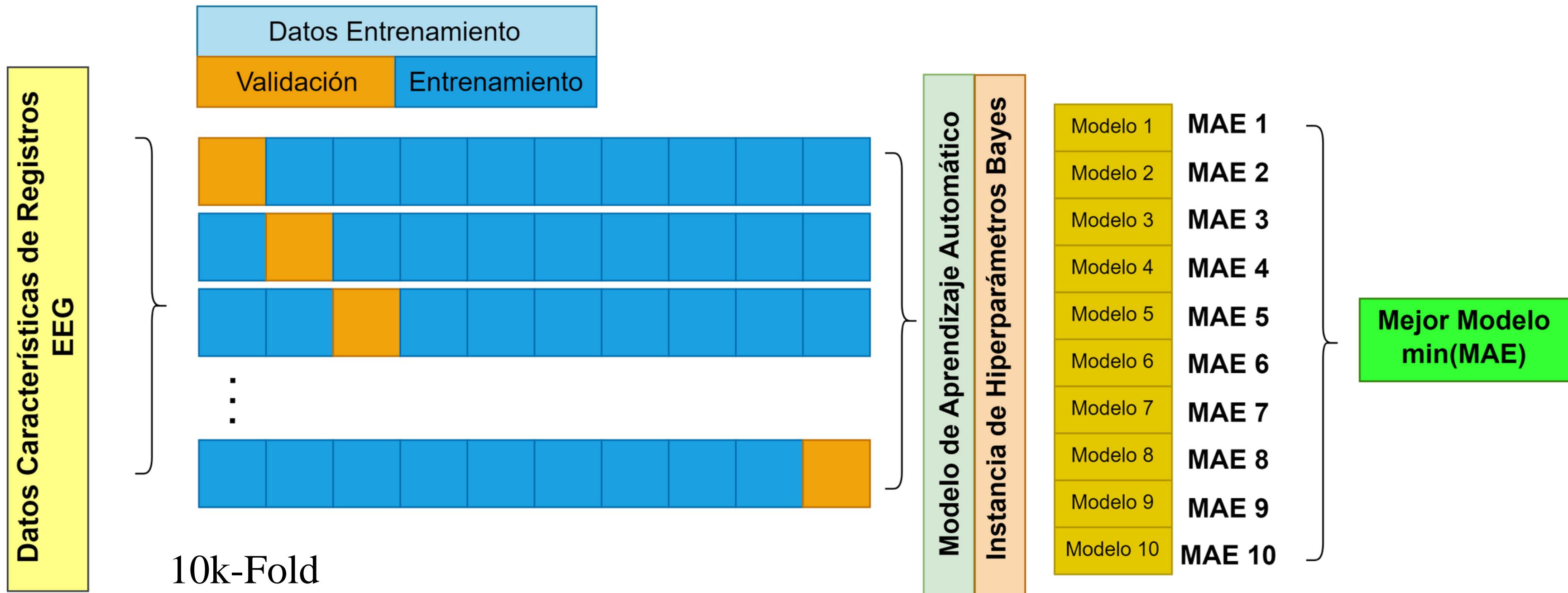
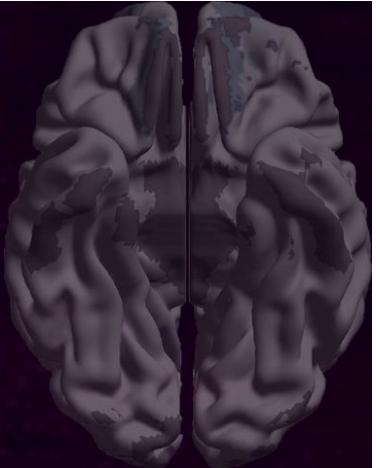
Búsqueda de Hiperparámetros



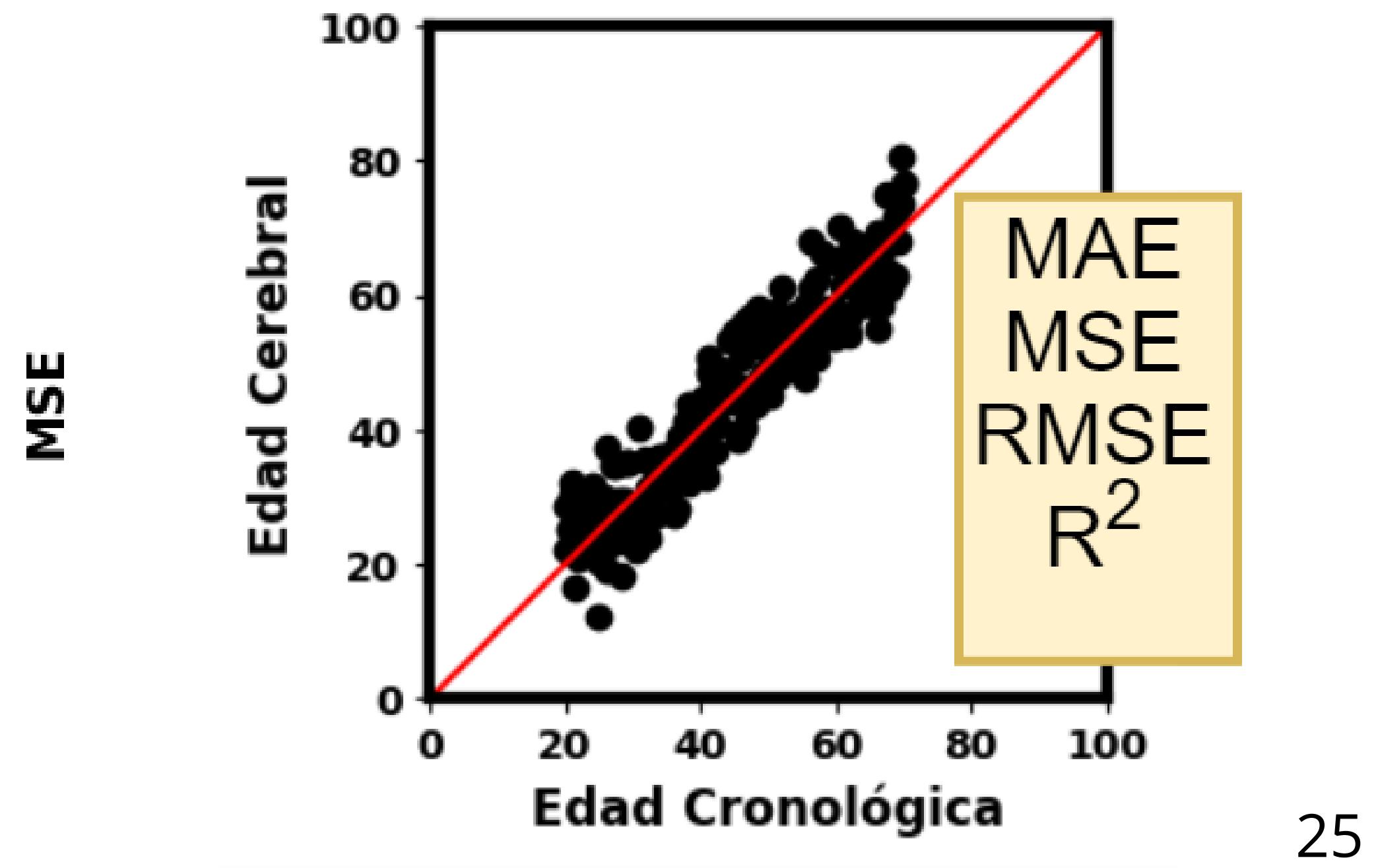
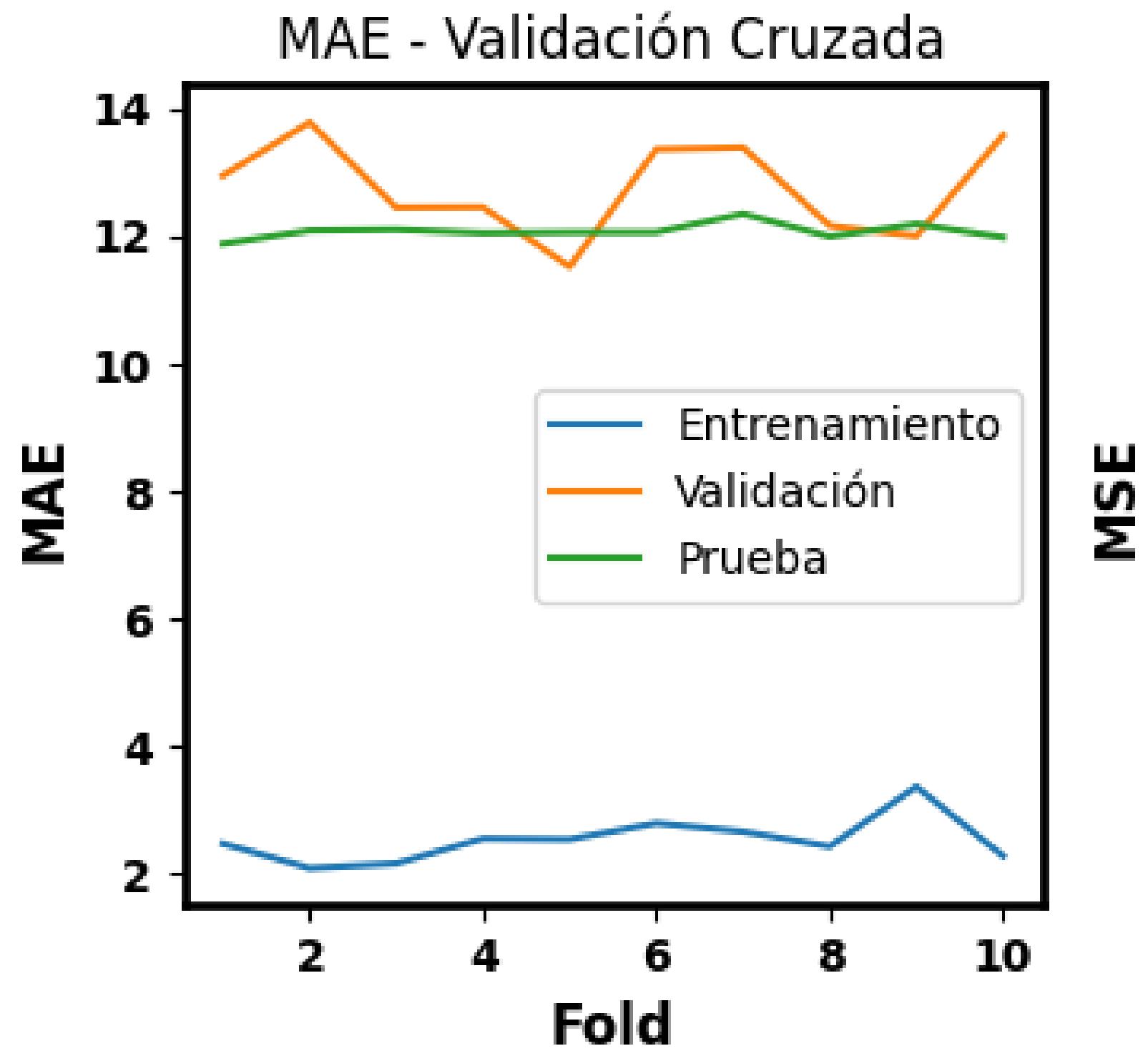
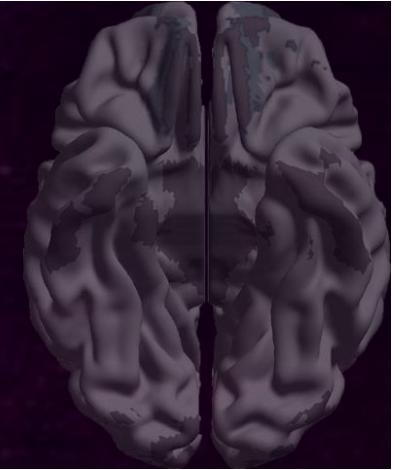
Selección de Hiperparámetros



Entrenamiento del modelo

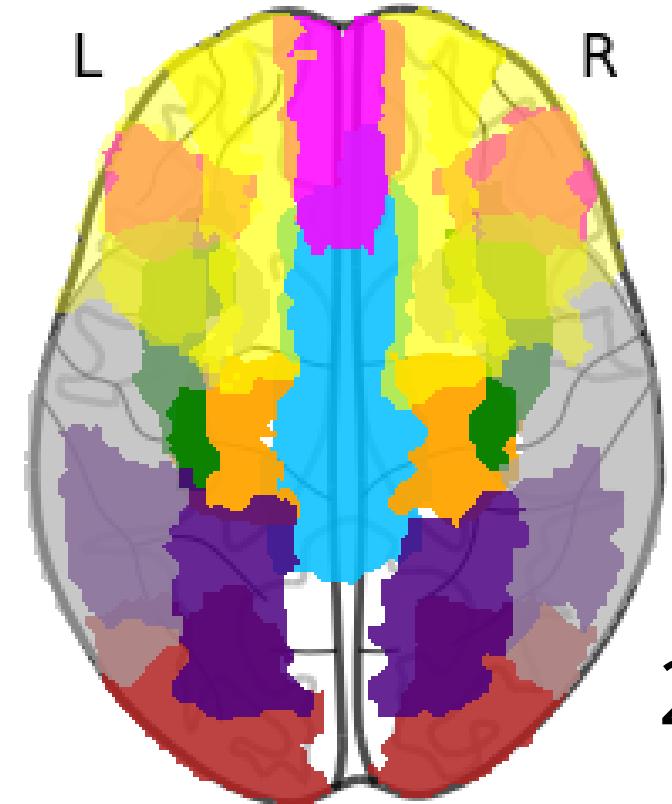
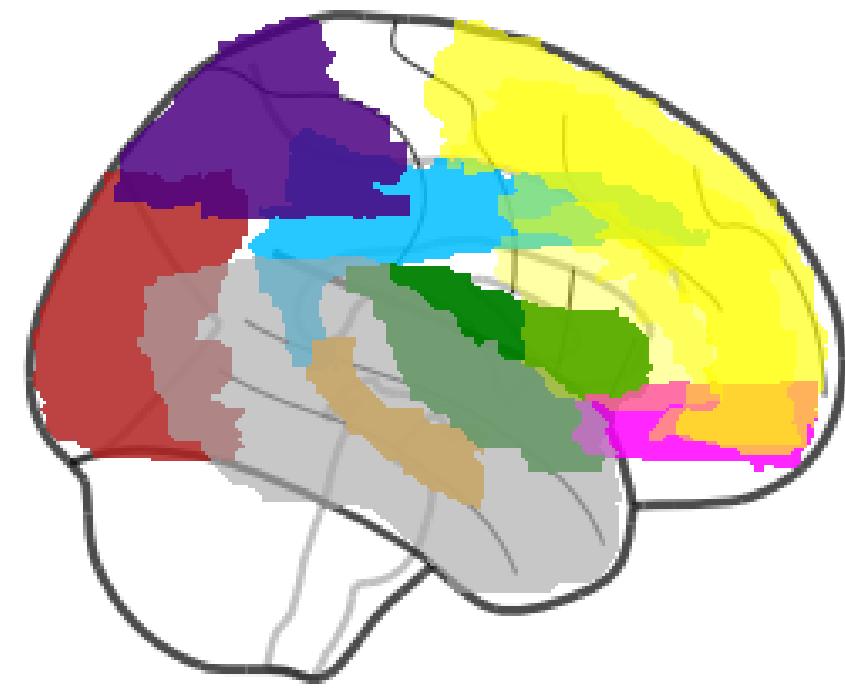
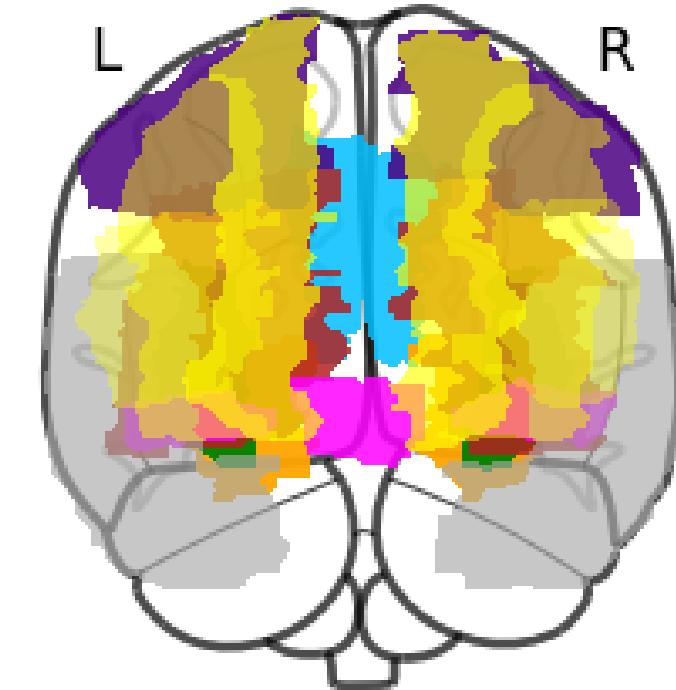
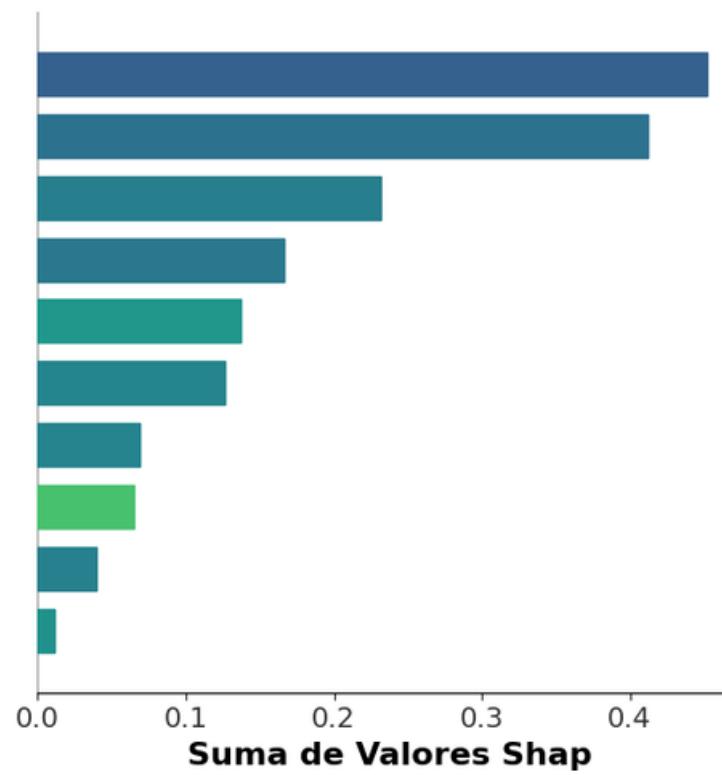
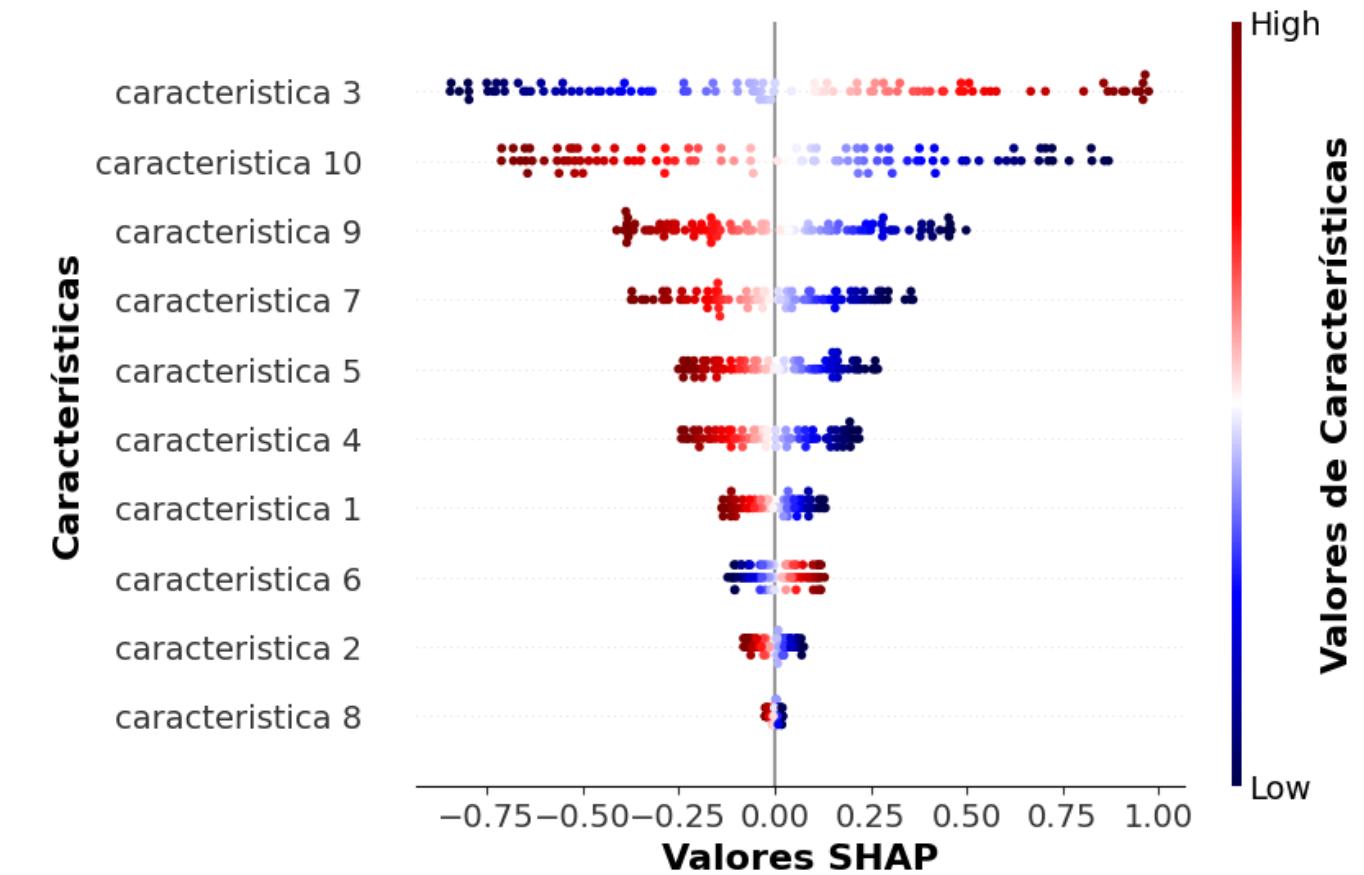
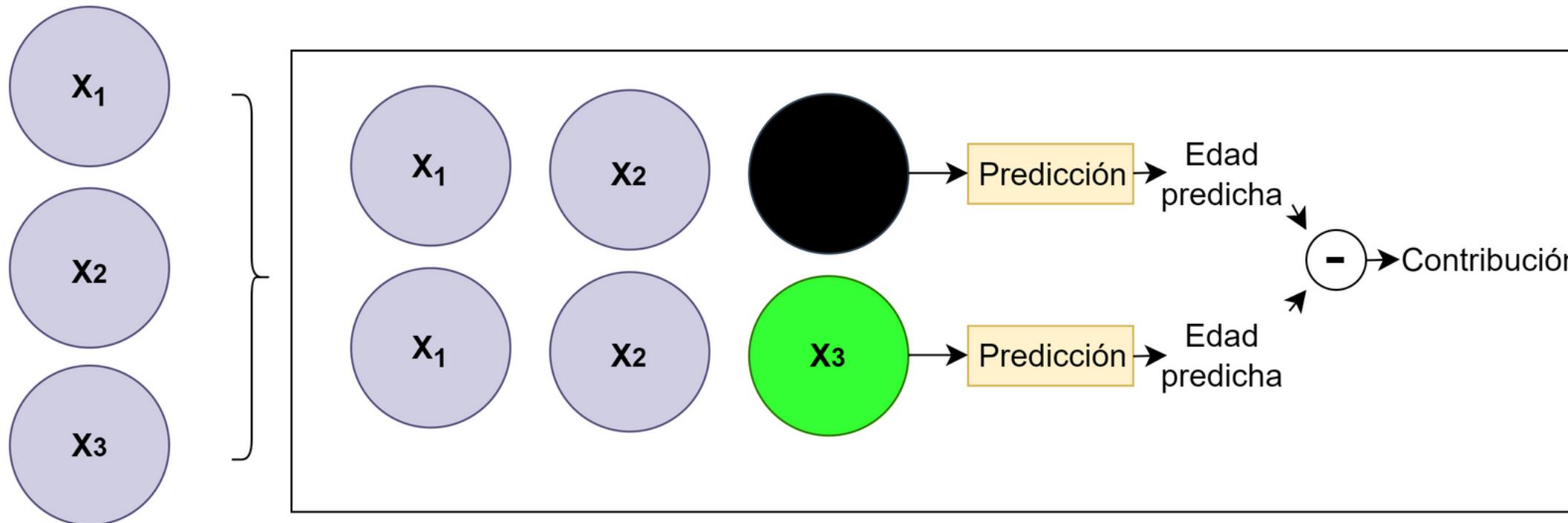


Entrenamiento del modelo



Interpretación del Modelo: SHAP

Característica EEG



Support Vector Regression

Función

Objetivo:

$$\min_{\mathbf{w}, b} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right)$$

Función de Predicción:

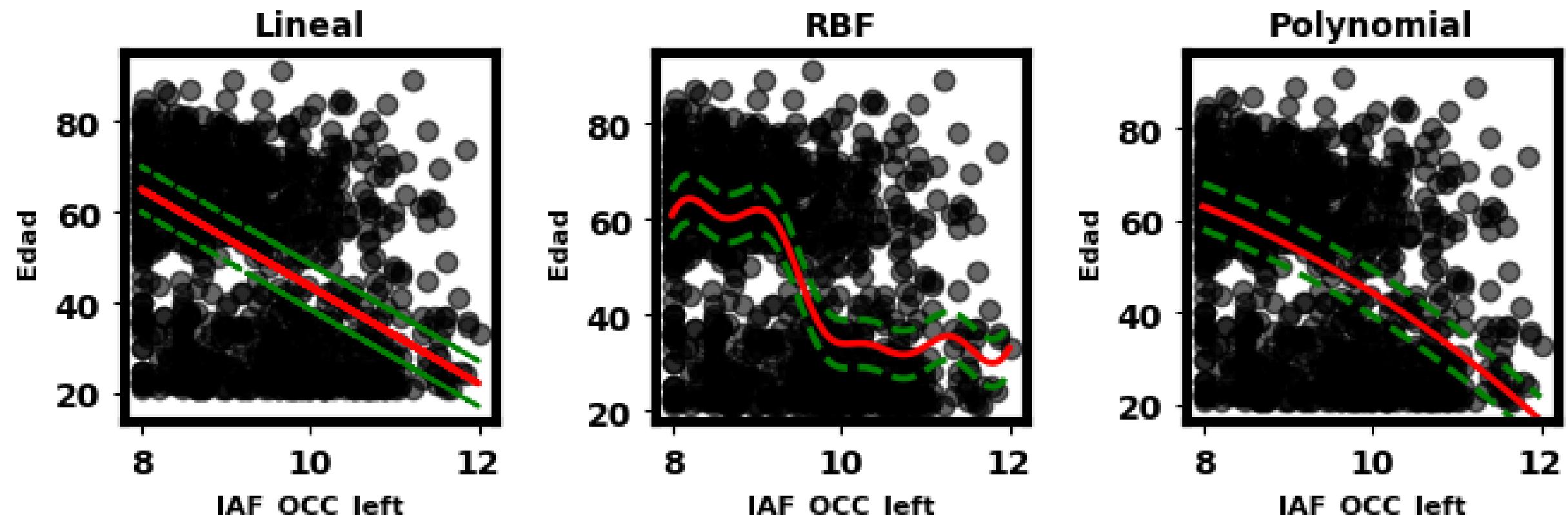
$$\hat{y} = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

Hiperparámetros:

C : 100

epsilon: 1

kernel: RBF



$$\begin{aligned} y_i - \mathbf{w}\mathbf{x}_i - b &\leq \epsilon + \xi_i & \text{para } i = 1, 2, \dots, n \\ -y_i + \mathbf{w}\mathbf{x}_i + b &\leq \epsilon + \xi_i^* & \text{para } i = 1, 2, \dots, n \\ \xi_i, \xi_i^* &\geq 0 & \text{para } i = 1, 2, \dots, n \end{aligned}$$

K vecinos más cercanos (KNN)

Función de Predicción:

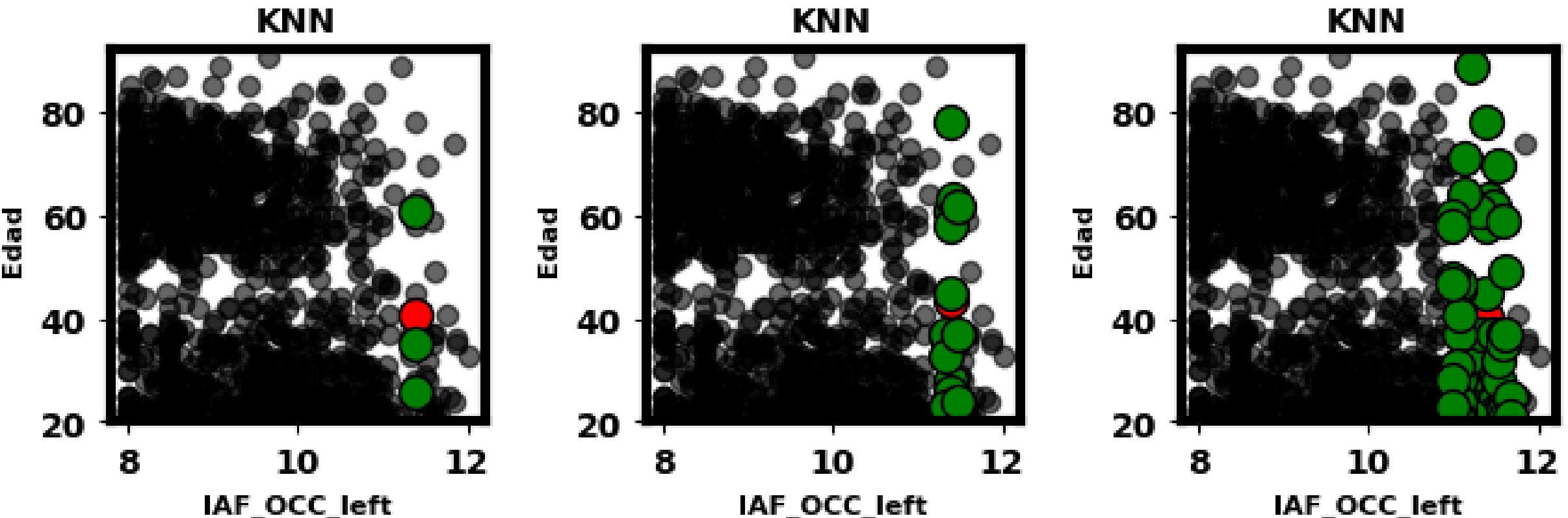
$$\hat{y} = \frac{1}{K} \sum_{i=1}^K y_i$$

Hiperparámetros:

K_vecinos: 14

Pesos: Distancia

Distancia_p: Manhattan



$$\text{Distancia Euclidiana}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2}$$

$$\text{Distancia Manhattan}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^d |(x_{ik} - x_{jk})|$$



Random Forest

Función

Objetivo:

$$E = \frac{1}{n} \sum_{i=1} l(y_i, \hat{y}_i)$$

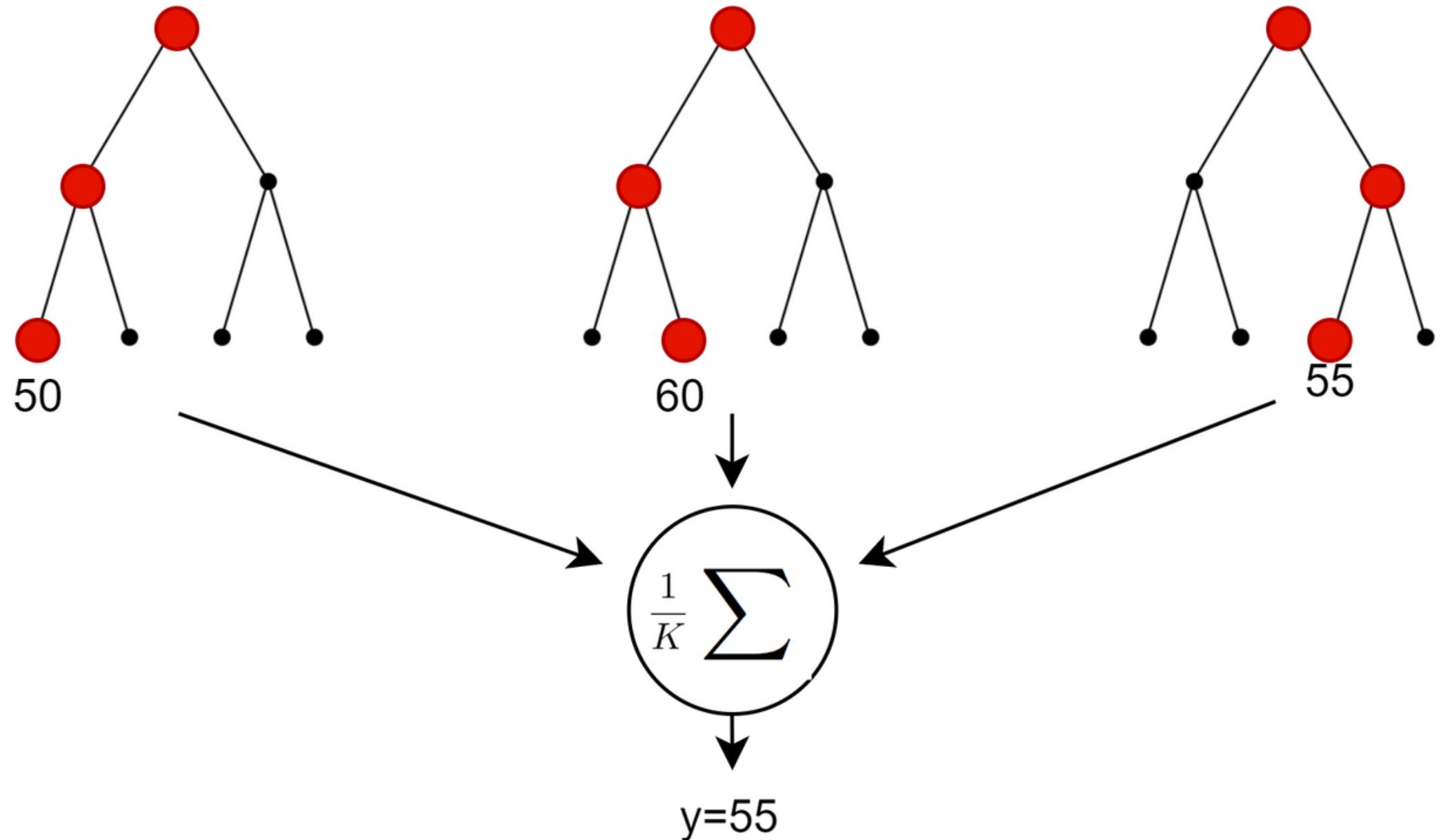
Función de Predicción:

$$\hat{y} = \frac{1}{K} \sum_{i=1}^K \hat{y}_i$$

Hiperparámetros:

Número de Arboles : 898

Profundidad máxima: 46



XGBoost y CatBoost

Función

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

Función de Predicción:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i)$$

Hiperparámetros XGboost:

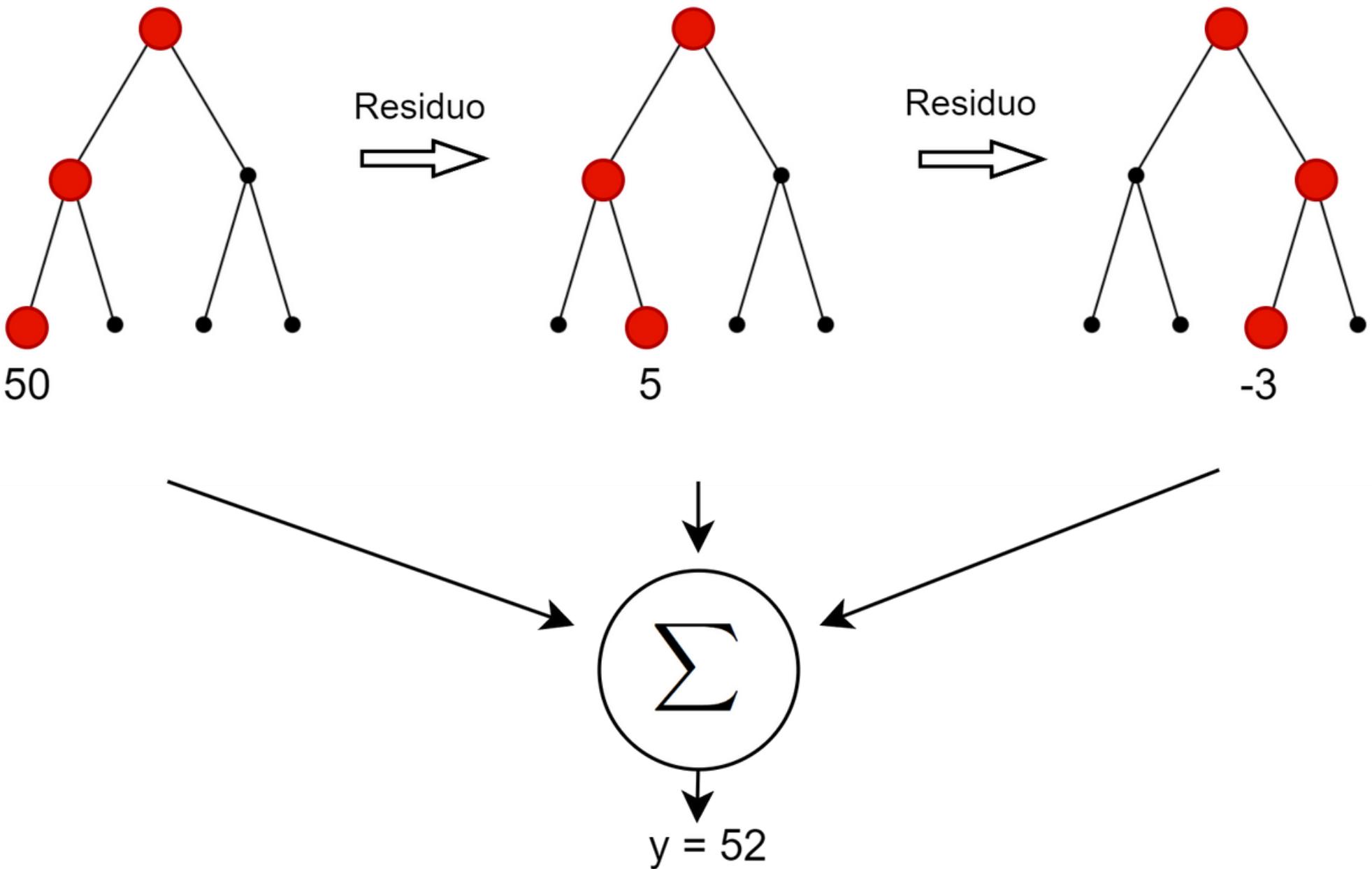
Número de Arboles : 550

Profundidad máxima: 7

Hiperparámetros Catboost:

Número de Arboles : 1334

Profundidad máxima: 7



Catboost: Ordered Boosting

Red Neuronal Multicapa (MLP)

Función

Objetivo:

$$L(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

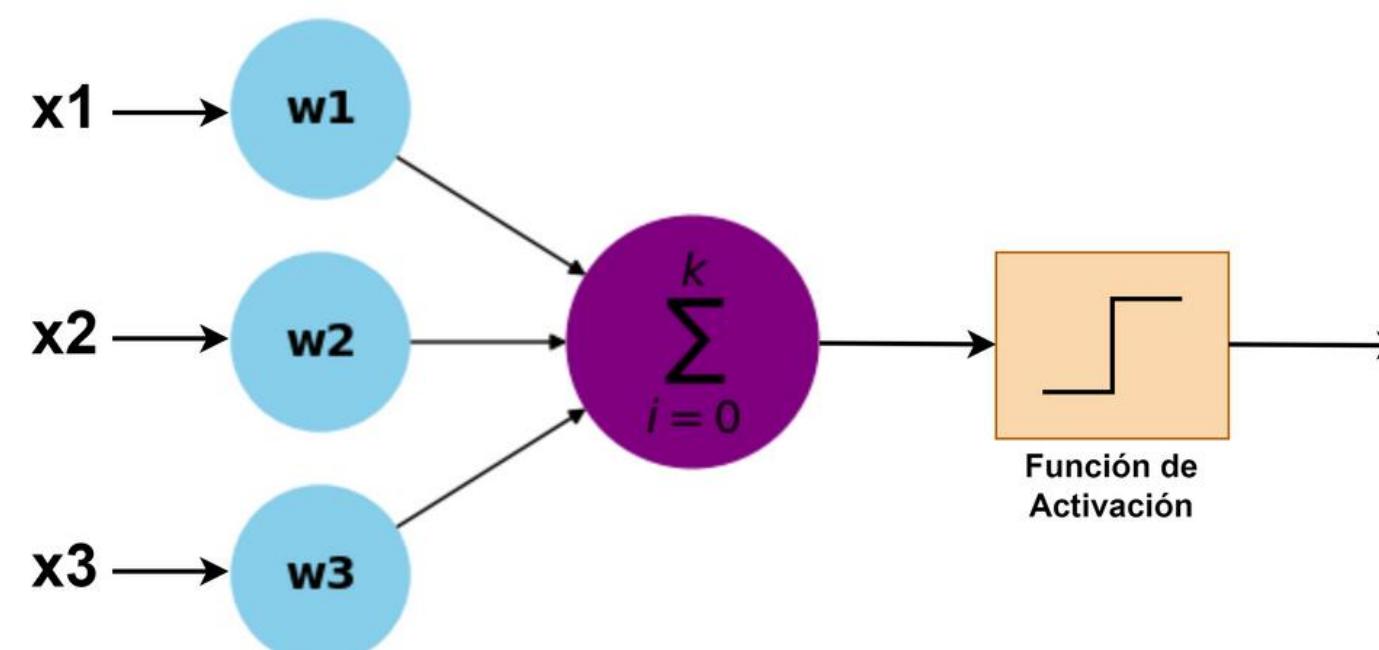
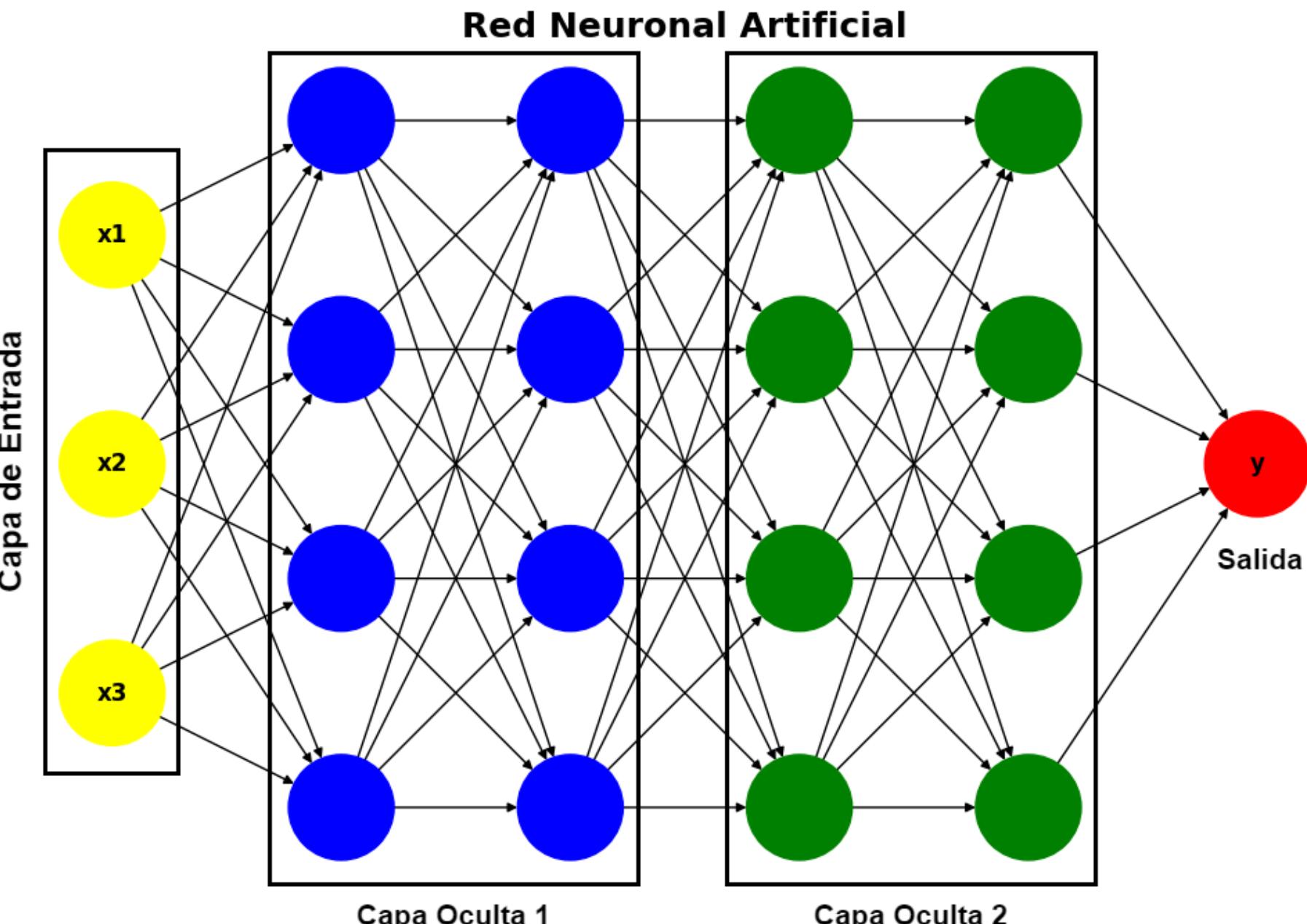
$$z_j = \sum_{j=1}^k w_{ij} x_i + b_j$$

Hiperparámetros:

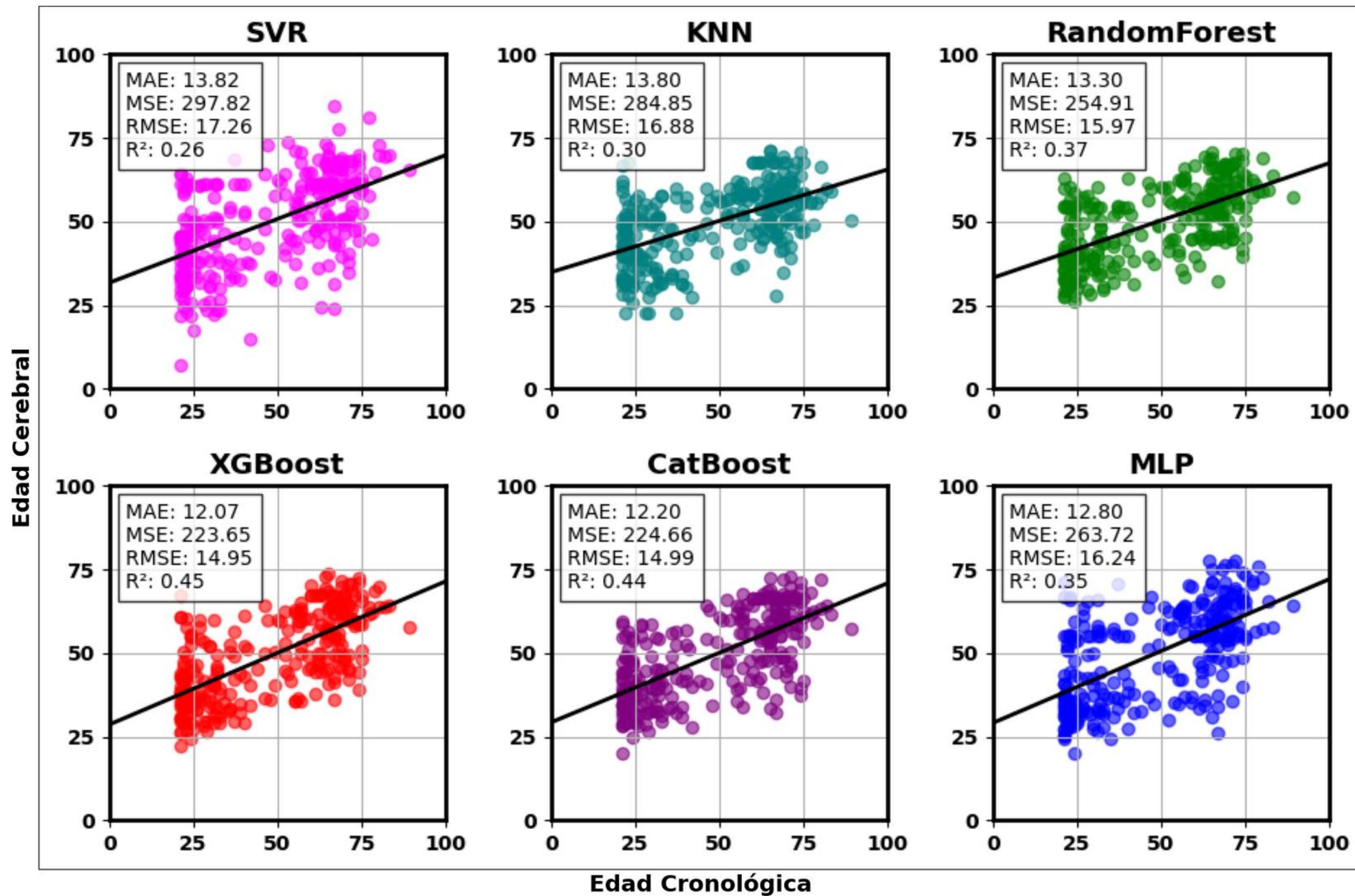
Capa 1 : 1584

Capa 2: 644

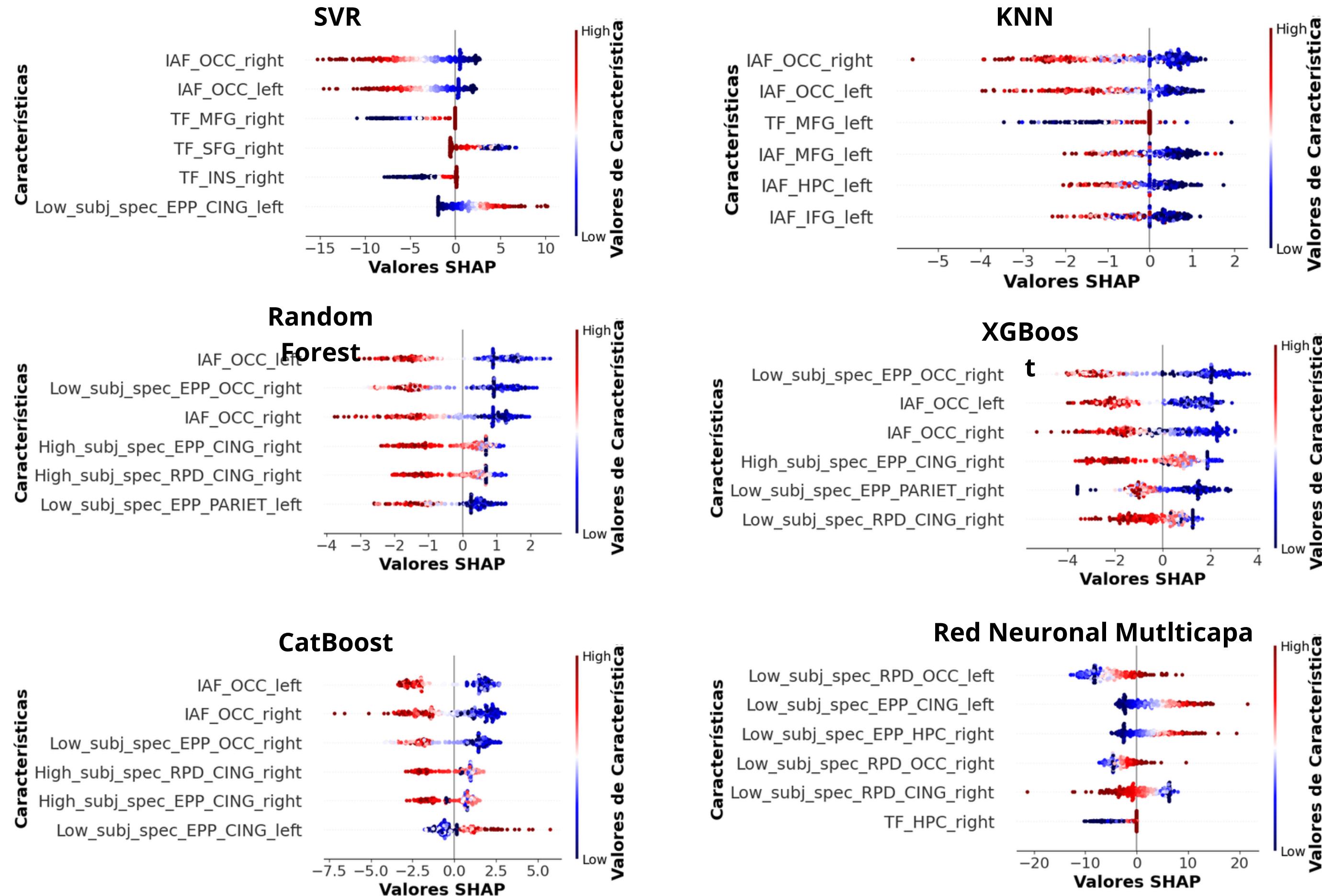
Activación: Logistic



Resultados

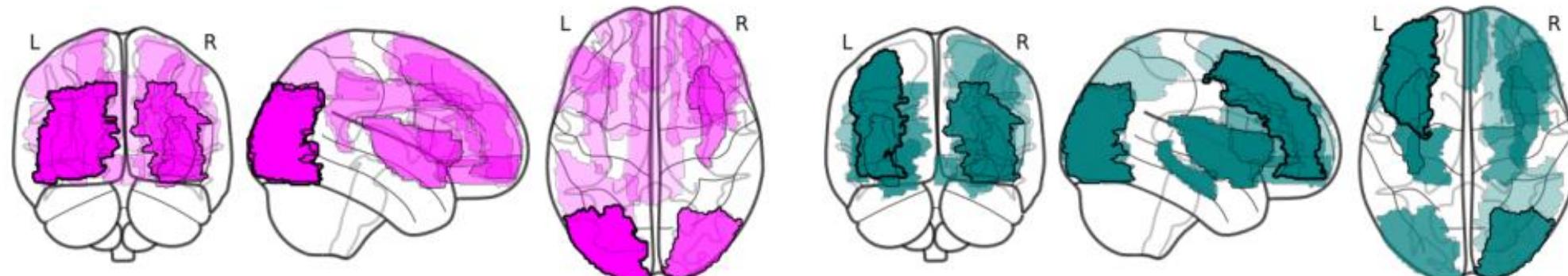


Valores SHAP

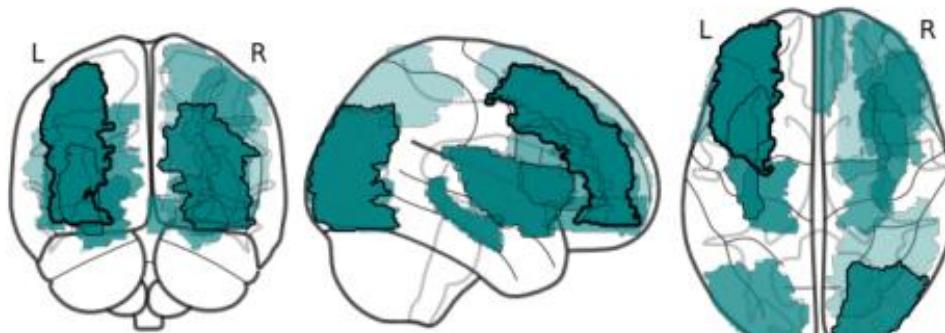


Importancia por Región

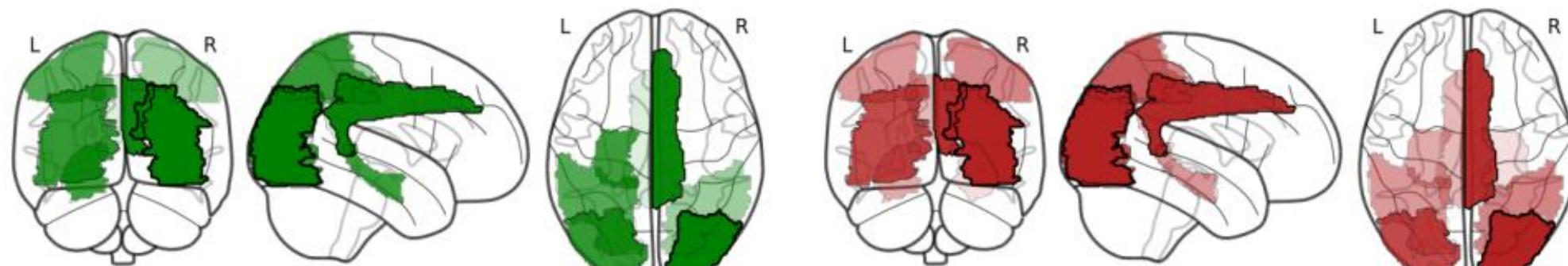
SVR



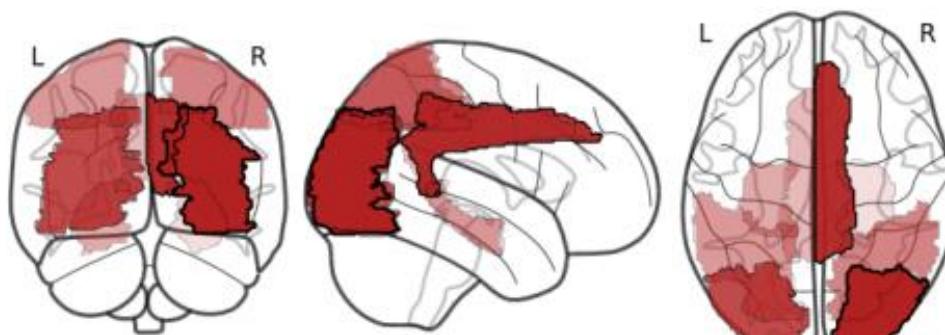
KNN



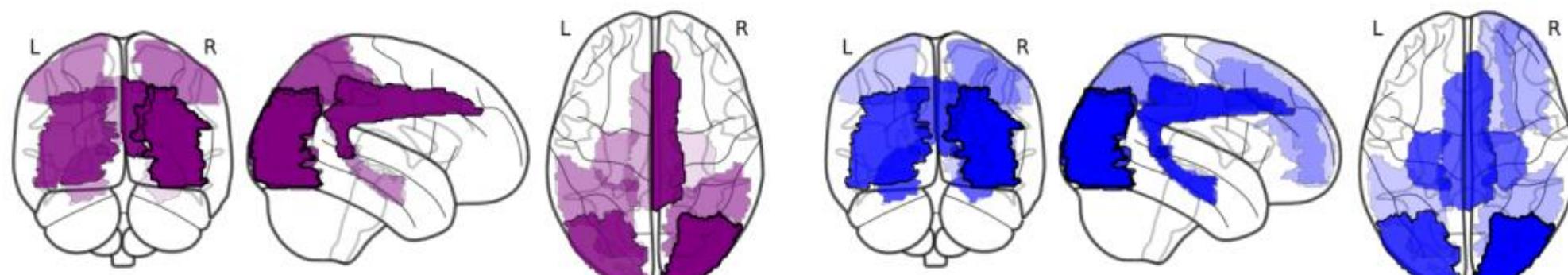
Random Forest



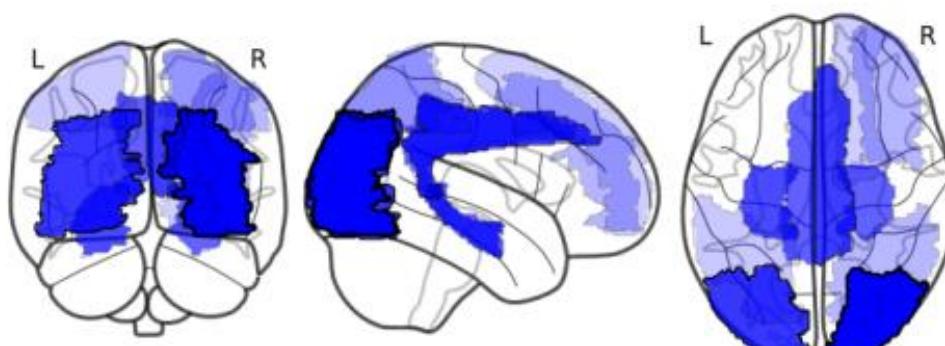
XGBoost



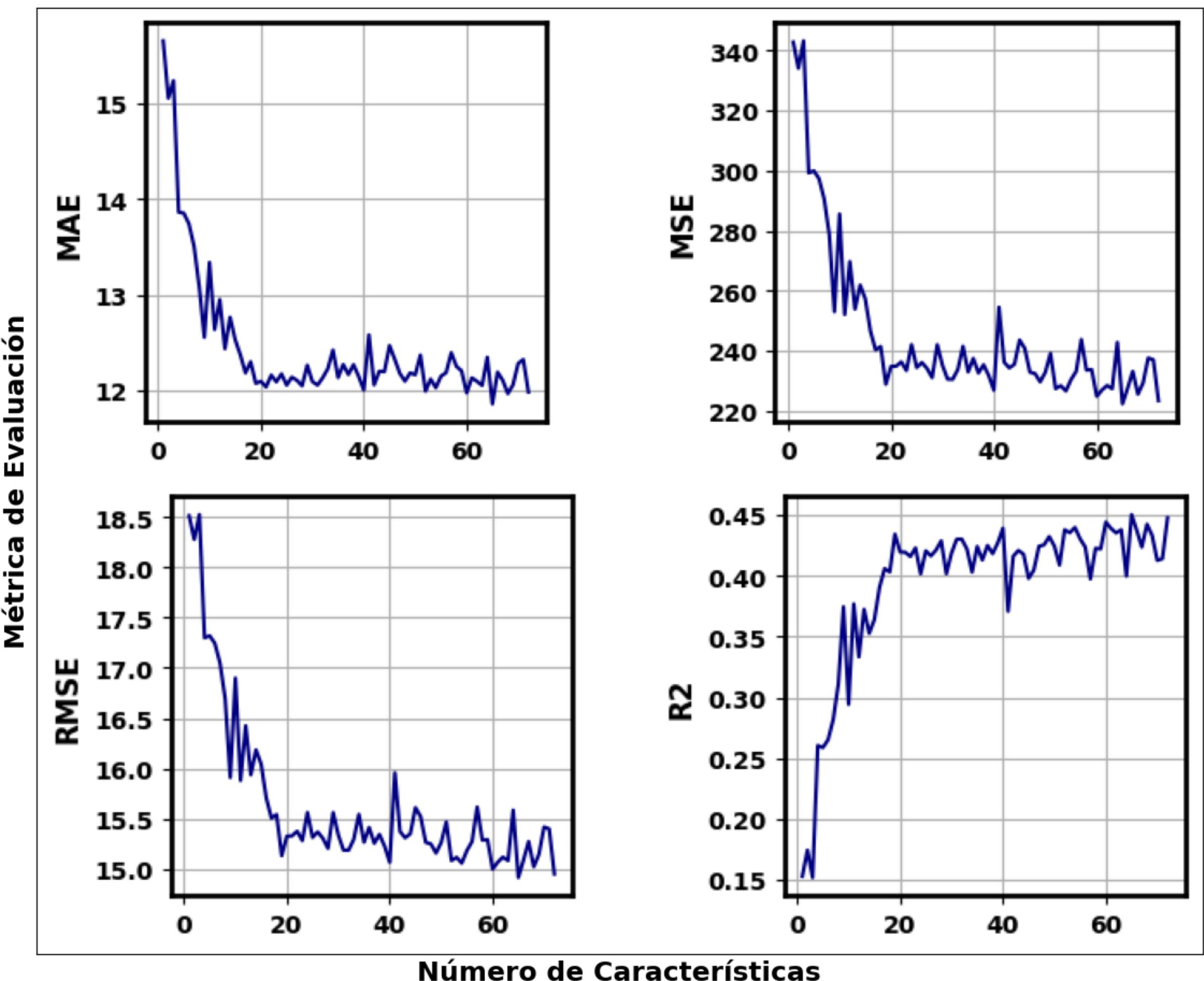
CatBoost



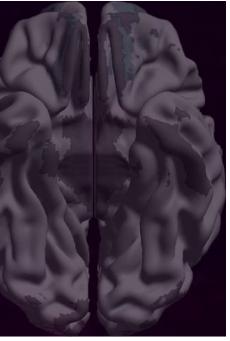
MLP



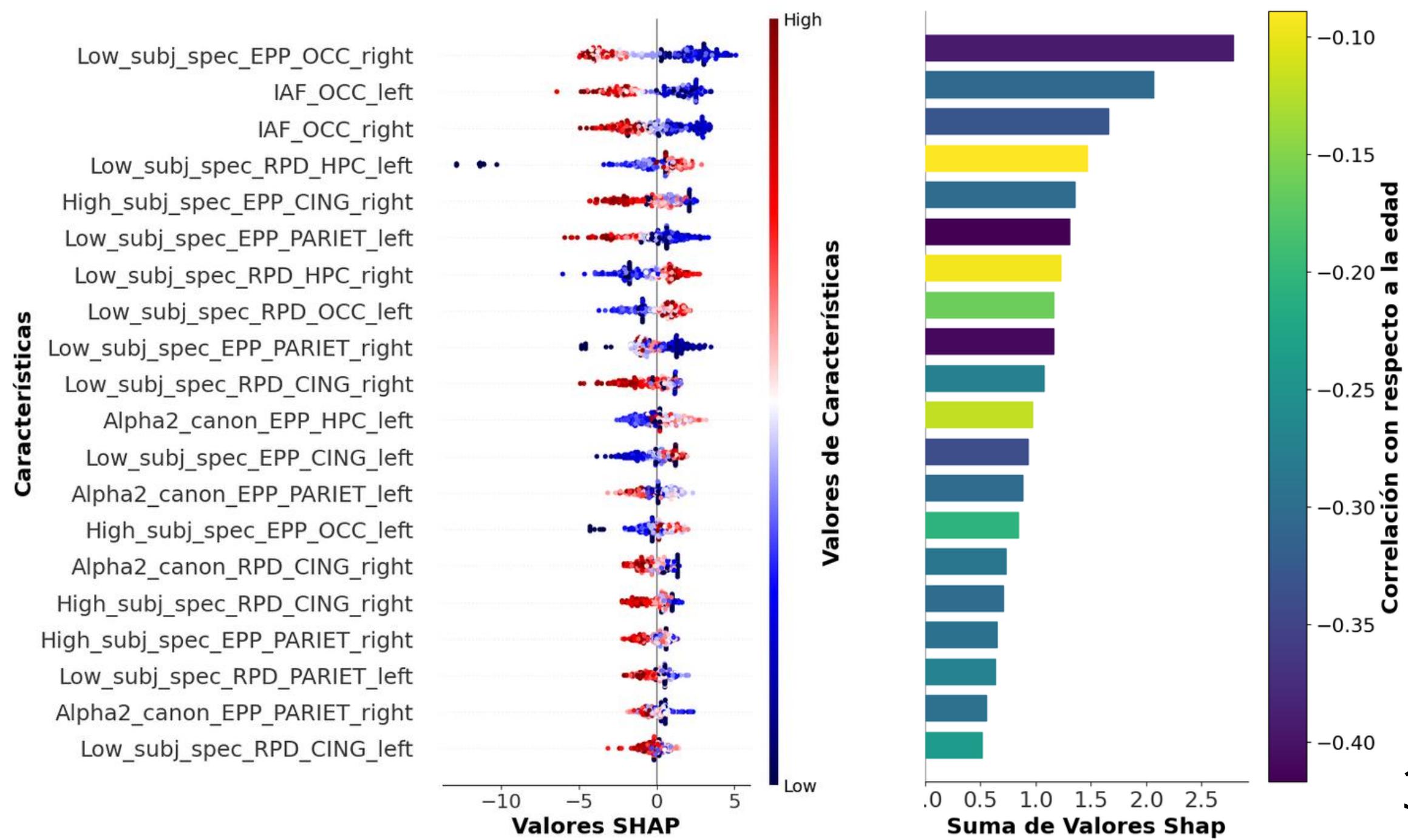
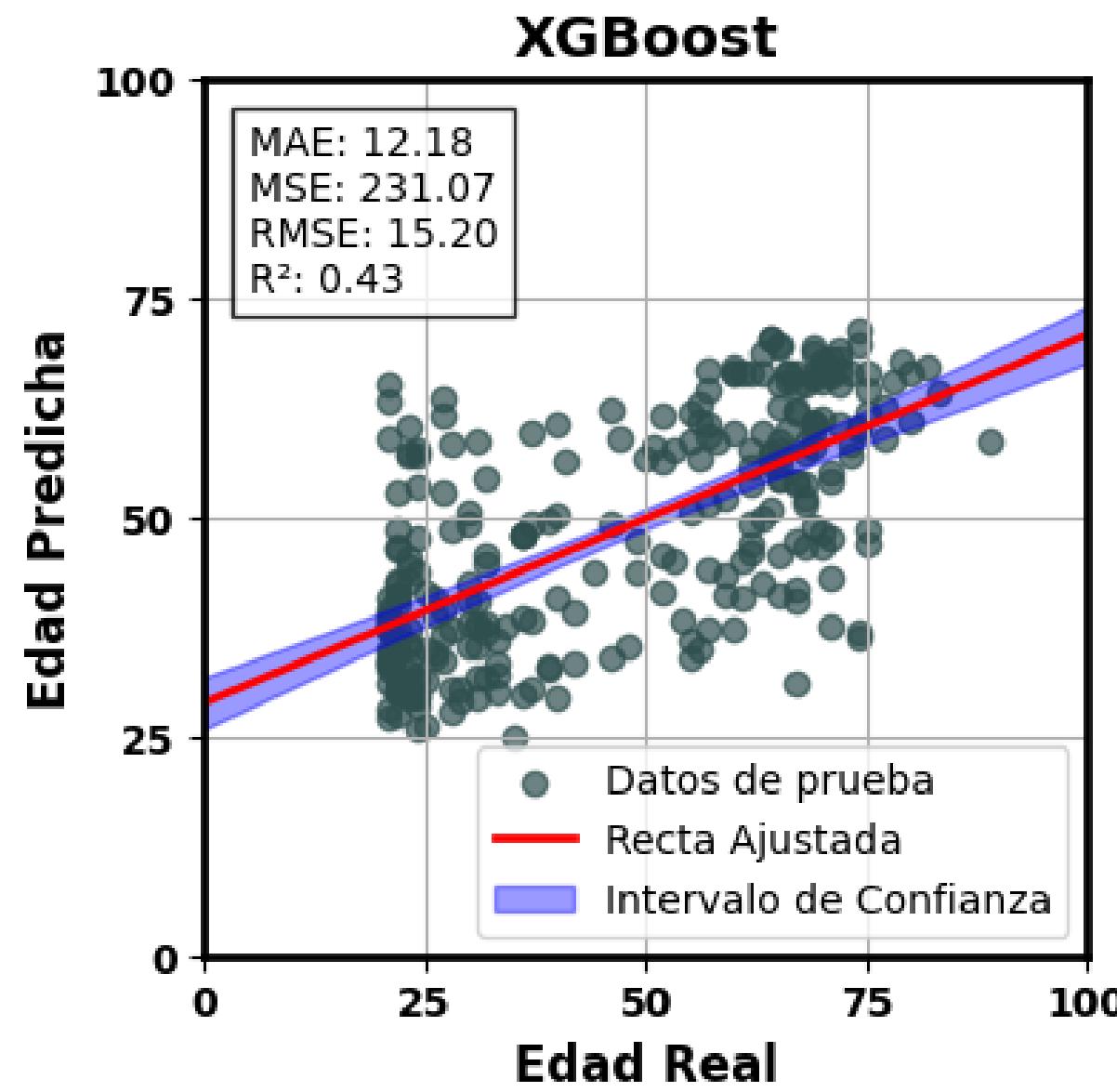
Evaluación incremental de características para modelo XGBoost



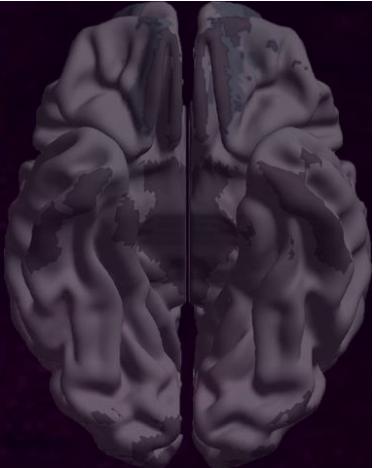
XGBoost



Modelo Evaluado para las 20 Mejores Características:

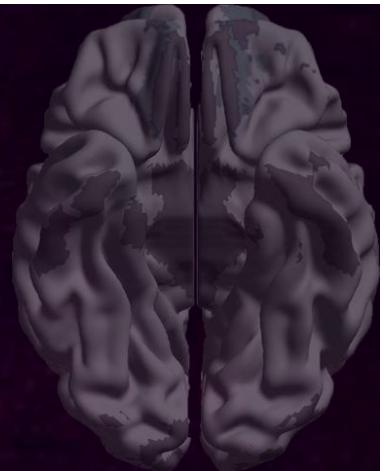


Conclusiones



- Los algoritmos de aprendizaje de máquina captan patrones complejos de actividad cerebral que no pueden apreciarse con estadística clásica.
- Todos los algoritmos son óptimos para realizar la estimación de edad cerebral, aunque unos son mejores que otros.
- La región occipital es la más influyente al momento de predecir la edad cerebral.

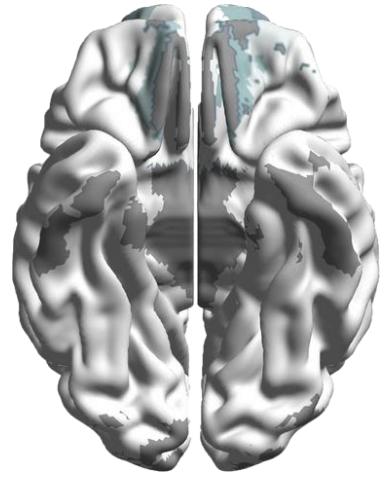
Trabajos Actuales



- Estimación de la edad biológica cerebral vs la edad cronológica utilizando un solo electrodo.
- Evaluación del modelo en Envejecimiento cerebral (Mayores de 40 años).
- Evaluación del modelo relacionando diversas patologías al envejecimiento cerebral.

Hay quienes luchan un dia y son buenos. Hay otros que luchan un año y son mejores. Hay quienes luchan muchos años y son muy buenos. Pero hay los que luchan toda la vida: esos son los imprescindibles.

- Bertolt Brecht



Muchas Gracias

